Coffalyser.NET analysis manual

Title	Coffalyser.NET analysis manual beta version
Status	Release candidate
Classification	Confidential
Versie	0.1

Version management

Version: 0.1
Status: Concept
Date: 22 July 2013
Classification: Confidential

Version history

Version	Date	Status	Auteur	Owner	Comments
0.1	09-03-12	Concept	Jordy Coffa	Jordy Coffa	Initial version

Foreword

This document contains the analysis manual for Coffalyser.NET. This document is created as a release candidate matching the first beta version (v120316.1250).

Amsterdam, 22 July 2013

Jordy Coffa

Title	Coffalyser.NET analysis manual beta version
Status	Release candidate
Classification	Confidential
Versie	0.1

Contents

1. Background
1.1 Introduction to MLPA normalization

2. Logging in
2.1 Organizations and user accounts
2.2 User Access
2.3 Connecting to a local database
2.4 Connecting to a server database

3. About the Coffalyser.NET sheet manager
3.1 What is the Coffalyser.NET sheet manager
3.2 Downloading new sheets
3.3 About products, lots and version
3.4 Viewing the available products, lots and versions
3.5 Adjusting a sheet lot
3.6 Adding a new “custom” product to the sheet manager

4. About CE devices
4.1 What are the CE devices?
4.2 Adding a new machine
4.3 Choosing the correct machine
4.4 Choosing the correct filter set
4.5 Changing the CE devices settings

5. About projects
5.1 What does a project contain?
5.2 Creating a new project
5.3 Project settings

6. About experiments
6.1 What does an experiment contain?
6.2 Creating a new experiment

6.3 Experiment settings
6.4 Setting the experiment type
6.5 Setting the channel contents
6.6 Setting the channel settings

7. About the fragment analysis
7.1 Importing the data files
7.2 Starting the fragment analysis.
7.3 About the fragment analysis quality scores.
7.4 Using the fragment results explorer

8. About the comparative analysis (copy number)
8.1 Setting up the comparative analysis
8.2 About the comparative analysis quality scores
8.3 About the comparative experiment results explorer
8.4 About the comparative sample results explorer

9. Methylation specific MLPA analysis
9.1 Introduction to MS-MLPA analysis
9.2 Setting up the MS-MLPA analysis
9.3 Comparative analysis experiment explorer with MS-Data
9.4 9.4 Comparative analysis sample explorer with MS-Data

10. FAQ
10.1 What is the different when the analysis method is set to RNA

11. References

1. Background

1.1 Introduction to MLPA normalization

MLPA kits generally contain about 40-50 oligo-nucleotide probes targeted to mainly the exonic regions of a single or multiple genes. The number of genes that each kit contains is dependent on the purpose of the designed kit. Each oligo-probe consists of two hemi-probes, which after denaturation of the sample DNA hybridize to adjacent sites of the target sequence during an overnight incubation. For each probe oligo-nucleotide in a MLPA kit there are about 600.000.000 copies present during the overnight incubation. An average MLPA reaction contains 60 ng of human DNA sample, which correlates to about 20.000 haploid genomes. This abundance of probes as compared to the sample DNA allows all target sequences in the sample to be covered. After the overnight hybridization adjacent hybridized hemi-probe oligo-nucleotides are then ligated using a ligase enzyme and the ligase cofactor NAD at a slightly lower temperature than the hybridization reaction (54 ℃ instead of 60 ℃). The ligase enzyme used, L igase-65, is heat-inactivated after the ligation reaction. Afterwards the non-ligated probe oligonucleotides do not have to be removed since the ionic conditions during the ligation reaction resemble those of an ordinary 1x PCR buffer. The PCR reaction can therefore be started directly after the ligation reaction by adding the PCR primers, polymerase and dNTPs. All ligated probes have identical end sequences, permitting simultaneous PCR amplification using only one primer pair. In the PCR reaction, one of the two primers is fluorescently labeled, enabling the detection and quantification of the probe products.
The different length of every probe in the MLPA kit then allows these products to be separated and measured using standard capillary fragment electrophoresis. The unique length of every probe in the probe mix is used to associate the detected signals back to the original probe sequences. These probe product measurements are proportional to the amount of the target sequences present in a sample but cannot simply be translated to copy numbers or methylation percentages. To make the data intelligible, data of a probe originating from an unknown sample needs to be compared with a reference sample. This reference sample is usually performed on a sample that has a normal (diploid) DNA copy number for all target sequences. In case the signal strengths of the probes are compared with those obtained from a reference DNA sample known to have two copies of the chromosome, the signals are expected to be 1.5 times the intensities of the respective probes from the reference if an extra copy is present. If only one

copy is present the proportion is expected to be 0.5. If the sample has two copies, the relative probe strengths are expected to be equal. In some circumstances reliable results can be obtained by comparing unknown samples to reference samples simply by visual assessment. This can be done best by overlaying two fragment profiles and comparing relative intensities of fragments (figure 1).

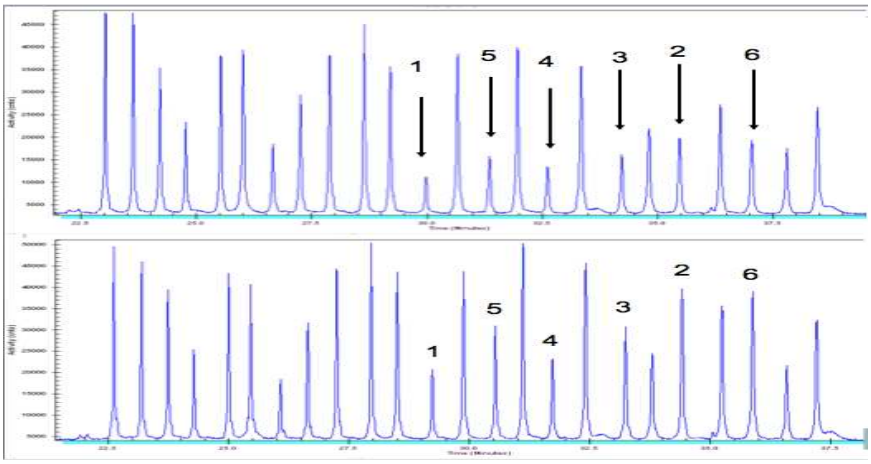


Figure 1 *MLPA fragment profile of a patient sample with Canavan disease (top) and that of a reference sample (bottom). Canavan disease is the result of a defect in the ASPA gene on chromosome 17p13. The fragment profile shows that the probe signals targeted to exon 1-6 of the ASPA gene have a 50% decrease as compared to the reference, which may be the result of a heterozygous deletion.*

It may however not be feasible to obtain reliable results out of such a visual comparison if:

1) The DNA quality of the samples and references is incomparable.
2) The MLPA kit contains probes targeted to a number of different genes or different chromosomal regions, resulting in complex fragment profiles
3) The data set is very large, making visual assessment very laborious.
4) The DNA was isolated tumor tissue, which often shows DNA profiles with altered reference probes

To make (complex) MLPA data easier understandable unknown and reference samples have to be brought on a common scale. This can be done by normalization, the division of multiple sets of data by a common variable in order to cancel out that variable's effect on the data. In MLPA kits,

so called reference probes are usually added, which may be used in multiple ways in order to comprise a common variable. Reference probe are usually targeted to chromosomal regions that are assumed to remain normal (diploid) in DNA of applicable samples. The results of data normalization are probe ratios, which display the balance of the measured signal intensities between sample and reference. In most MLPA studies, comparing the calculated MLPA probe ratios to a set of arbitrary borders is used to recognize gains and losses (González, 2008). Probe ratios of below 0.7 or above 1.3 are for instance regarded as indicative of a heterozygous deletion (copy number change from two to one) or duplication (copy number change from two to three), respectively. A delta value of 0.3 is a commonly accepted empirically derived threshold value for genetic dosage quotient analysis (Bunyan et al. 2004). Since chromosomal aberrations often span larger regions, ordering probe data by Map View locations(NCBI, Map view version 36) results in clustering
of probes targeting the same region. Aberrations can be recognized more easily this way and probes targeting the same region may confirm each other's result.
This criterion alone may often not provide the conclusive results required for diagnosing disease. MLPA probes all have their own characteristics and the level of increase or decrease that a probe ratio displays that was targeted to a region that contains a heterozygous gain or loss, may differ for each probe. Interpretation of normalized data may even be more complicated due to shifts in ratios caused by sample-to-sample variation such as: dissimilarities in PCR efficiency and size to signal sloping. Other reasons for fluctuations in probe ratios may be: poor amplification, misinterpretation of an artifact peak/band as a true probe signal, incorrect interpretation of stutter patterns or artifact peaks, contamination, mislabeling or data entry errors (Bonin et al., 2004). To make result interpretation more reliable our software combines effect-size statistics and statistical interference allowing users to evaluate the magnitude of each probe ratio in combination with it's significance in the population. The significance of each ratio can be estimated by the quality of the performed normalization, which can be assessed two factors: the robustness of the normalization factor and the reproducibility of the sample reactions.
In this document we show the features and integrated analysis methods of our novel MLPA analysis software called Coffalyser.NET. Our software uses an analysis strategy that can adapt to fit the researcher objectives while considering both the biological context and the technical limitations of the overall study. We use statistical parameters appropriate to the situation, and apply the most robust normalization method based on the biology and quality of the data. Most information required for the analysis is extracted directly from the MRC-Holland database, producer of the MLPA technology,

needing only little user input about the experimental design to define an optimal analysis strategy. In the next chapters we explain how we can use this software to analyze a MLPA experiment, create experiment overview reports, sample reports and chart and how we can make sense of the found results.

2. Logging in

2.1 Organizations and user accounts

Our software uses a SQL client–server database model to store all project/experiment-related data. The client-server model has one main application (server) that deals with one or several slave applications (clients). Clients may communicate to a server over the network, allowing data sharing within and even beyond their institutions. Even though this system may provide great convenience e.g. for people who are working on a single project but are working on different locations, both client and server may also reside in the same system. Having both client and server on the same system has some advances over running both separately: the database is better protected and both client and server will always have the same version number. In case an older client will try to connect to a server that has a newer version number, the client needs to be updated first. A client does not share any of its resources, but requests a server's content or service function. Clients therefore initiate communication sessions with servers that await incoming requests. When a new client is installed on a computer it will implement a discovery protocol in order to search for a server by means of broadcasting. The server application will then answer with its dynamic address that resolves any issues with dynamic IP addresses.

2.2 User Access

In addition to serving as a common data archive, the database provides user authentication, robust and scalable data management, and flexible archive capabilities via the utilities provided within Software. Our database model acts in accordance with a simple legal system, linking users to one or multiple organizations. Each user receives a certain role within each organization to which certain right are linked. These rights may for instance include denial of access to certain data but may also be used to deny access to certain parts of the program. These same levels may also be applied on project level. Projects will have project administrators and project members. The initial project creators will also be the project administrators who are responsible for user management of that project.

2.3 Connecting to a local database

After installing Coffalyser.NET on a computer in a standalone configuration, the first screen you will find each time you start the program in the user login screen. If you did not install Coffalyser.NET yet, please find installation instruction in the Coffalyser.NET installation manual. To login to your local database make sure that the configuration is set to Single PC / Standalone, next fill in the user name of the account you have created during the installation and the password and click on Login.

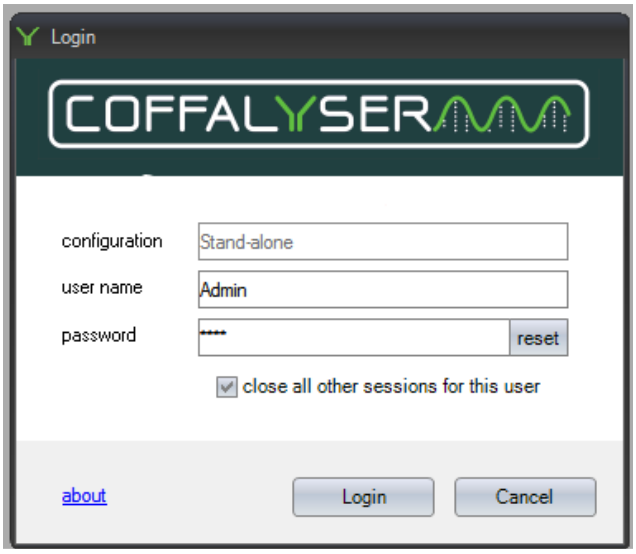


Figure 2 *Coffalyser.NET login form for single PC, standalone configurations*

If your login fails, please check the configuration settings of your database. To change or check your configuration, click on the "Windows Start button" then navigate to "All programs" and search the entree "Coffalyser.NET". From the Coffalyser.NET program menu, click on "Configure Coffalyser.NET"

2.4 Connecting to a server database

If you have installed Coffalyser.NET on a computer in a client / server configuration, the login procedure will be similar to that described at 2.3, the configuration settings however will differ. Make sure that the configuration at

the login screen is set to “Client in Multi PC / Networked” as noted in figure 3.

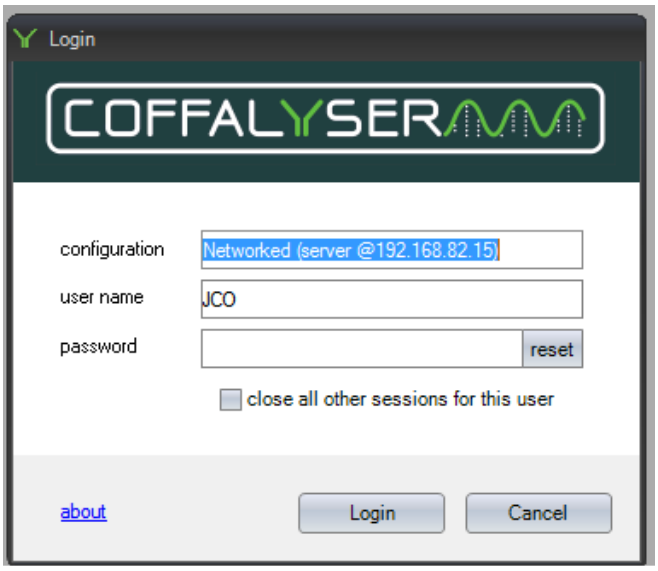


Figure 3 *Coffalyser.NET login form for clients in multi PC / networked environments.*

To get instruction on how to set up a computer as a Coffalyser.NET server, I want to refer to the “Coffalyser.NET installation manual”. The server name should be the IP-address of the computer where the Coffalyser.NET server is installed and the port number should be “1231”. If you don’t know the IP-address of your server computer, start command prompt on the server computer (click “start menu”, click “run”, type “cmd”; press enter) and type in command prompt “Ipconfig /all”. Your IP address should turn up in the list.

3. About the Coffalyser.NET sheet manager

3.1 What is the Coffalyser.NET sheet manager

Coffalyser.NET is equipped with MLPA sheet manager software, allowing users to obtain information about commercial MLPA kits and size markers directly from the MRC-Holland database. Next to this, the sheet manager also allows users to create custom MLPA mixes. The sheet manager software can be used to check if updates to any of the MLPA mixes are available. The sheet manager can further carry out automatic checks for updates at the frequency you choose, or it can be used to make manual checks whenever you wish. It can display scheduled update checks and can work completely in the background if you choose. With just one click, you can check to see if there are new versions of the program, or updated MLPA mix sheets. If updates are available, you can download them quickly and easily. In case some MLPA mixes are already in use, users may choose to hold on to both the older version and updated versions of the mix or replace the older version.

3.2 Downloading new sheets

Each Coffalyser.NET version starts with an empty database, containing only information about the created organizations, users within the database and standard capillary electrophoresis threshold needed for fragment analysis. To obtain the necessary information for data analysis, right click on “Sheet manager” in the database explorer and select “Download updates” from the right click menu, as indicated by the arrow in figure 4.

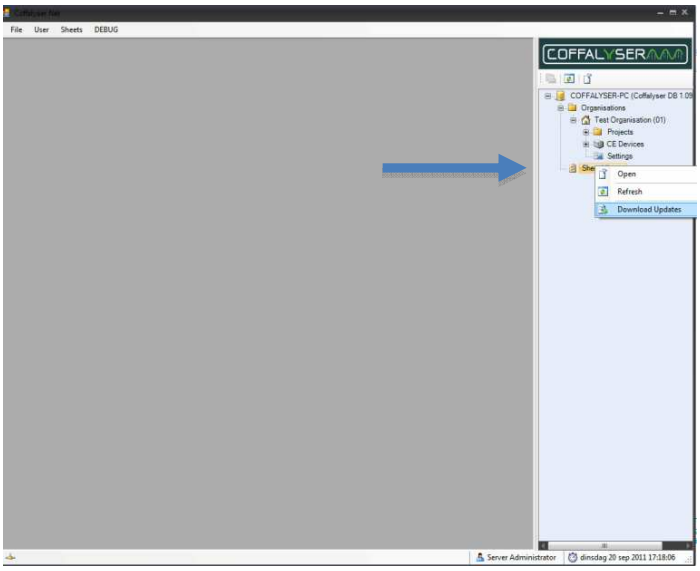


Figure 4 *Coffalyser.NET start screen and location to download sheet updates.*

Next you will see the download form, click on “Start Update” to download the latest MLPA sheet. In case you are using restricted products from MRC-Holland, please email to: Coffalyser@mlpa.com, to receive the download code. This code can be entered in the box which appears by clicking on the "*Add Code*" button. Adding codes to the database will enable certain restricted lots to be used, which are normally not downloaded (figure 5). Directly after you will see an update window, showing the progress of the current update (figure 6). When the update is finished you will receive a message declaring that the download was successful or unsuccessful. If the download was unsuccessful please check if your internet connection is active, or if you firewall is blocking internet access of the program.

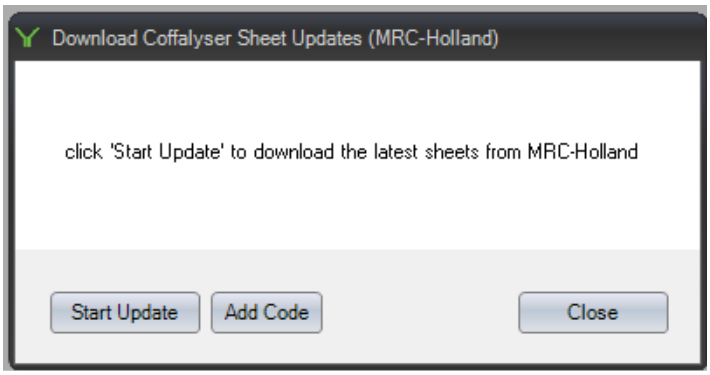

Download Coffalyser Sheet Updates (MRC-Holland)
click 'Start Update' to download the latest sheets from MRC-Holland
Start Update
Add Code
Close

Figure 5 *Coffalyser.NET start update screen.*

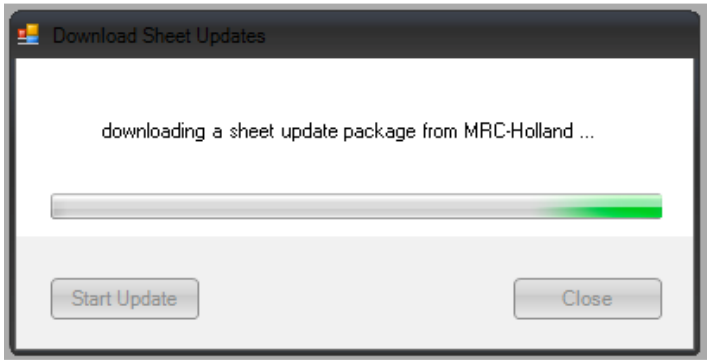


Figure 6 *Coffalyser.NET start screen and location to download sheet updates.*

3.3 About products, lots and active sheets

To circumvent problems with designated products and lots, and to make sure you will be able to find the correct product and lot you are using, the sheet manager has viewing and editing capabilities to compare your product description with the available products in the sheet manager. Each product developed by MRC-Holland has a P-number for copy number products, a M-number for products that can be used for detection of methylation status or a R-number indicating products that can be used for quantification of RNA sequences. To view the list with all available products right click on the on “Sheet library” in the database explorer and select “Open” from the right click menu (figure 4). Next the currently active Coffalyser Work Sheets will open. By selecting “*Add*” from the right click context menu you can create a new

Coffalyser Work Sheet based on an existing MRC-Holland MLPA product or create an empty work sheet starting from scratch (figure 7).

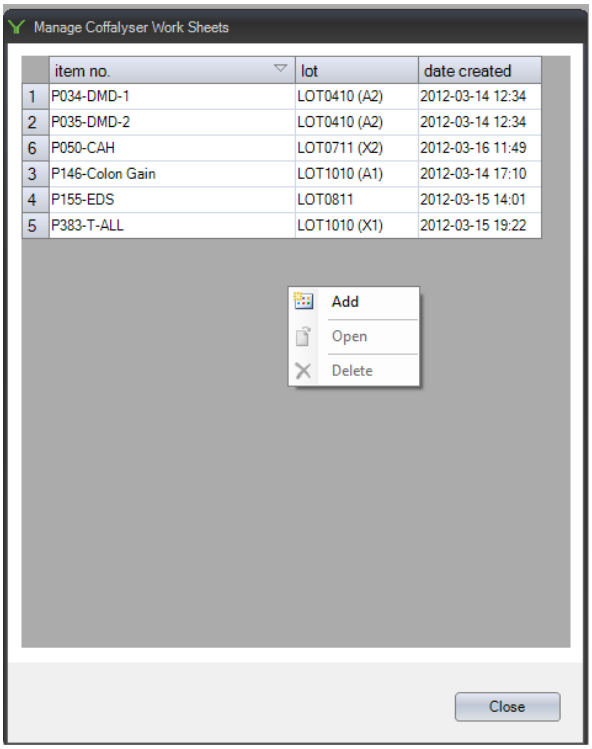


Figure 7 *Coffalyser.NET sheet library product overview window.*

3.4 Viewing the available products, lots and active sheets

To view a product, first add it to the Coffalyser Work Sheet library by using the right click context menu. After you add a product the Coffalyser Work Sheet Editor will automatically open allowing you to make the necessary changes or view if the content are as expected. At the moment when you add a Coffalyser work sheet of any existing product you basically make a copy of the original worksheets of MRC-Holland. This copy can then be adjusted to your wishes, the originals however never change.

To view the content of a product lot version, right click on a sheet and select open. This will open the details properties related to that version (figure 14). You can find who created this MLPA mix, who modified it, to which organization this version is related, what type of control fragments were added to this mix, what the default analysis method is and at which date this version was created in the Coffalyser.NET database. By clicking on the tab called “probes” you can view what probes are present in this MLPA mix and obtain information about these probes and their related target sequences (figure 8 & 9).

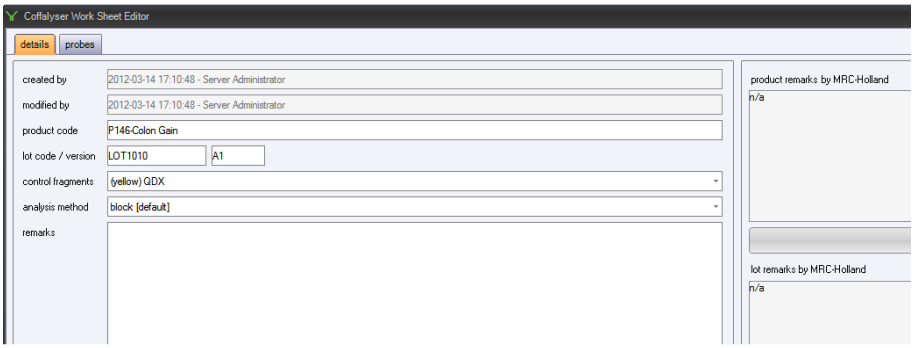


Figure 8 *Coffalyser.NET sheet library product lot version details window.*

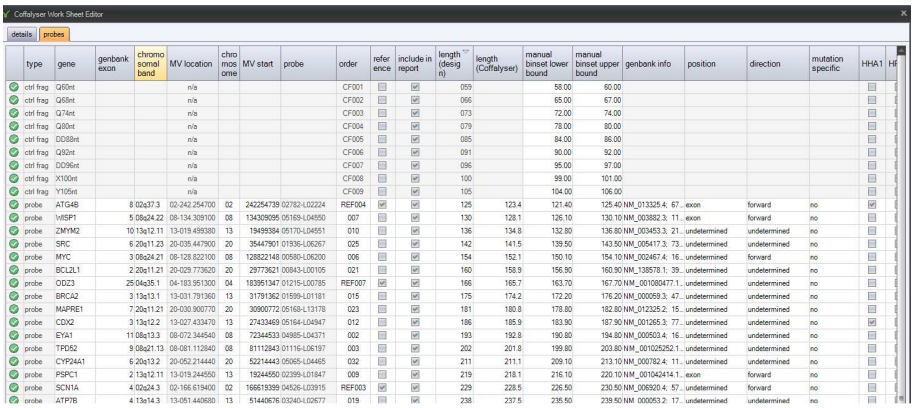


Figure 9 *Coffalyser.NET sheet library product lot version probes overview window.*

Title	Coffalyser.NET analysis manual beta version
Status	Release candidate
Classification	Confidential
Versie	0.1

3.5 Adjusting a sheet lot

You can now adjust to the organization to which this version belongs, the control mix that is present in this mix and the default analysis method. For more details about these fields, please check the chapter about control fragments and analysis settings in this manual.
The second tab; “probes”, can be used to adjust the separate details of each probe in the mix, use a single click in each cell to adjust each field. Most fields will automatically be checked if they contain valid values. Use the right click menu to add or remove probes, select “add” from the menu followed by the selection of the number of probes you wish to add. Please note, that after adding a series of probes all fields need to be filled in and validated. Right click on any probe and select “Validate Sheet Data” from the right click menu, to check if all added data can be used by the software. In case a probe has invalid information the related problems can be viewed by double clicking on the warning icon.
You may also change the control mix that is related to the probe mix by either changing the "*control fragments*" type fragments on the first page or by changing its in the probe editing grid in the right mouse click context menu. In the Coffalyser.NET 7 different control mixes are recognized, these being:

- (brown): contains the Q-fragments for DNA concentration check, Q92nt peak for ligation control, DD88nt & DD96nt for denaturation control, X100nt and Y105nt for gender check.
- (orange) Q92nt: only contains the Q92nt peak for ligation control probe
- (pink) QD: contains the Q-fragments for DNA concentration check, Q92nt peak for ligation control and the DD88nt & DD96nt for denaturation control for control of contamination with DNA, not for concentration estimations.
- (purple) MQD (mouse): equal to the (pink) QD but for mouse DNA
- (red) Q-fragments:): contain only the Q-fragments, this control mix is usually added to RNA products
- (yellow) QDX: an older version of the control mix (brown), contains the same fragment lengths but the DD88nt is less sensitive for higher salt concentrations than its equivalent in control mix (brown).
- (blue) BQDX (BIG): equal to (yellow) QDX but with adapted concentrations for (BIG) MLPA mixes

Please note that control mixes when using MLPA mixes bought from MRC-Holland are already pre-set to the proper control mix and should thus not be edited.

3.6 Adding a new “custom” product to the sheet manager

You may also add new “custom” product to the database. These are products that are made from scratch in your own laboratory based on synthetic or cloned oligo-nucleotides. To add a new product, start at the sheet library product overview window, right click anywhere in the window and then select “Add product” from the right click menu. This will open product lot details window allowing you to adjust the organization this sheet belongs to, the control fragments this mix contains and the default analysis method. The second tab, “probes” can then be used to add the probes to this mix and add the necessary details about these probes.

4. About CE devices

4.1 What are the CE devices?

CE devices within Coffalyser.NET are considered to be the capillary electrophoresis devices you are using to separate the MLPA products. Coffalyser.NET fragment analysis settings are specific for that machine and optimized settings for each known device are provided by defaults. Since detection of fluorescent units occurs on arbitrary scales and the measured intensities may also differ even from device to device (even of the same type) peak detection settings may often require empiric optimization. Each organization can therefore create a CE devices within the software that can be related to an actual device in your laboratory. When such a device is created, it will be loaded with the default fragment analysis properties provided by MRC-Holland. These settings will suffice in most cases, but in some cases manual optimization of these settings is required

4.2 Adding a new machine

Our software is compatible with binary data files produced by all major capillary electrophoresis systems including: ABIF files (*.FSA, *.AB1, *.ABI) produced by Applied Biosystems devices, SCF and RSD files produced by Megabace™ systems (Amersham™) and SCF and ESD files produced by CEQ systems (Beckman™). Before any data import can be performed we first need to define what machines are being used in your laboratory. Right click in the Coffalyser.NET database exploration window on the “CE Devices” and select “Add CE Device from the right click menu.

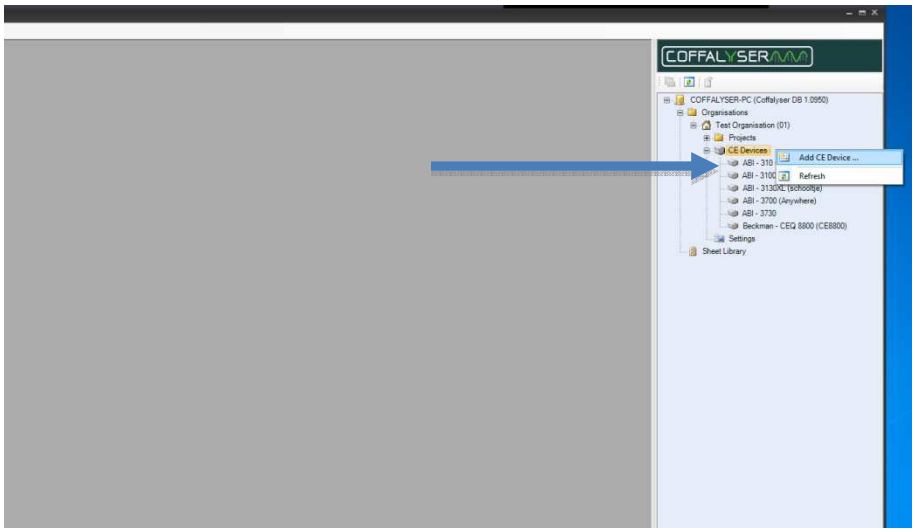


Figure 10 *Adding a new CE device to Coffalyser.NET.*

4.3 Choosing the correct machine

After selecting “Add device” a new window will open allowing you to define which capillary electrophoresis device your are using (figure 11). Next to “CE device”, choose the machine you wish to use form the dropdown list. If you are unsure what machine you are using please contact your provider.

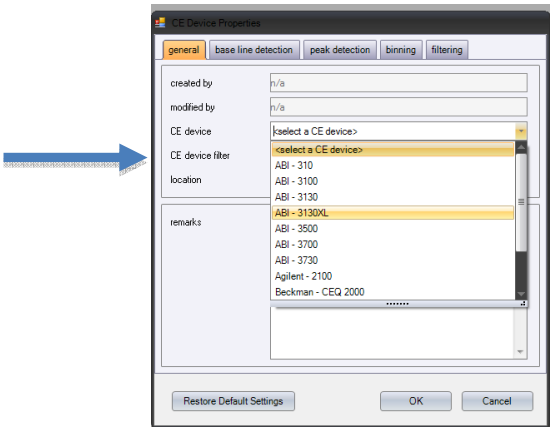


Figure 11 *Choosing your machine type.*

4.4 Choosing the correct filter set

After the correct machine was selected you should select the correct filter set that matches the chemistry you are using (figure 12). The filter sets defines what fluorescent dyes certain channels in the machine recognize. Table 1 contains the most common filter sets used by ABI. If for instance you are using a FAM label for the probes and you are using LIZ for a size marker, then you need to select filter set G5.

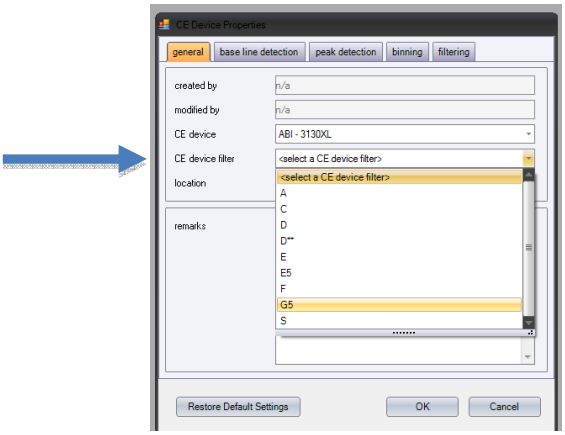


Figure 12 *Choosing your filter set.*

Table 1 *Filter sets used by ABI capillary devices.*

Dye Set	Filter Set	Blue	Green	Yellow	Red	Orange
DS-29	A	5-FAM™	JOE™	TAMRA™	ROX™	
DS-34	C	6-FAM™	TET™	HEX™	TAMRA™	
DS-30	D	6-FAM™	HEX™	NED™	ROX™	
DS-31	D**	6-FAM™	VIC®	NED™	ROX™	
DS-02	E5	dR110	dR6G	dTAMRA™	dROX™	LIZ®
DS-32	F	5-FAM™	JOE™	NED™	ROX™	
DS-33	G5	6-FAM™	VIC®	NED™	PET®	LIZ®

4.5 Changing the CE devices settings

By going through the different tabs of the CE Device properties window, you will be able to change the different device specific analysis settings. There are four types of settings that you may change, which are: baseline settings, peak detection settings, binning settings and filter settings. When you are working in a specific organization, CE device settings will be applied for all users that are working in that organization. In that case you need to be a administrative user to be able to change the CE device settings.

4.5.1 Baseline settings

When performing detection of fluorescence in capillary electrophoresis devices it is sometimes the case that spectra can be contaminated by fluorescence. Baseline curvature and offset are generally caused by the sample itself and little can be designed in an instrument to avoid these interferences (Nancy T. Kawai, 2000). Non-specific fluorescence or background auto fluorescence should be subtracted from the fluorescence

obtained from the probe products to obtain the relative fluorescence as a result of the incorporation of the fluorophore. The baseline wander of the fluorescence signals may cause problems in the detection of peaks and should be removed before starting peak detection. Our software corrects for this baseline by applying two times a median signal filter on the raw signals.
First, the signals of the first 200 (default is 80 for the size marker channel; see figure 19 "*Baseline moving median point marker*" and "*Baseline moving median points probes rough*") data points of each dye channel were extracted and its median was calculated. Then for every 200 subsequent data points till the end of the data stream, the same procedure was carried out. These median values are then subtracted from the signal of the original data stream to remove the baseline wander, resulting in baseline 1. For size marker channel no further correction is necessary since not much baseline wandering or shoulder peaks are expected.
For probe channels, this corrected baseline 1 is then fed as input for a filter that calculates the median signal over every 50 subsequent data points (figure 19 "*Baseline moving median points probes fine*"). Alternatively an advanced secondary baseline can be used which follows the baseline more accurately, as described at the fragment analysis settings chapter. These median values are then subtracted from all the signals that are below 300 RFU (see figure 19 "*Baseline maximum signal for correction fine*") on baseline 1, resulting in baseline 2. This second baseline is often necessary due to the relatively short distance between the peaks that derive from probe products with only a few nucleotides difference. By applying this second baseline correction solely on the signals that are in the lower range of detection, even peaks that reside close to each other may reside back to zero-signal, without subtracting too much fluorescence that originates from the probe products. Program administrators can modulate the default baseline correction settings, and also may store different defaults for each used capillary system.
In general it is not recommended to adjust the baseline settings, however if one may notice that certain peaks are not being detected it may be necessary to change these settings (figure 13).

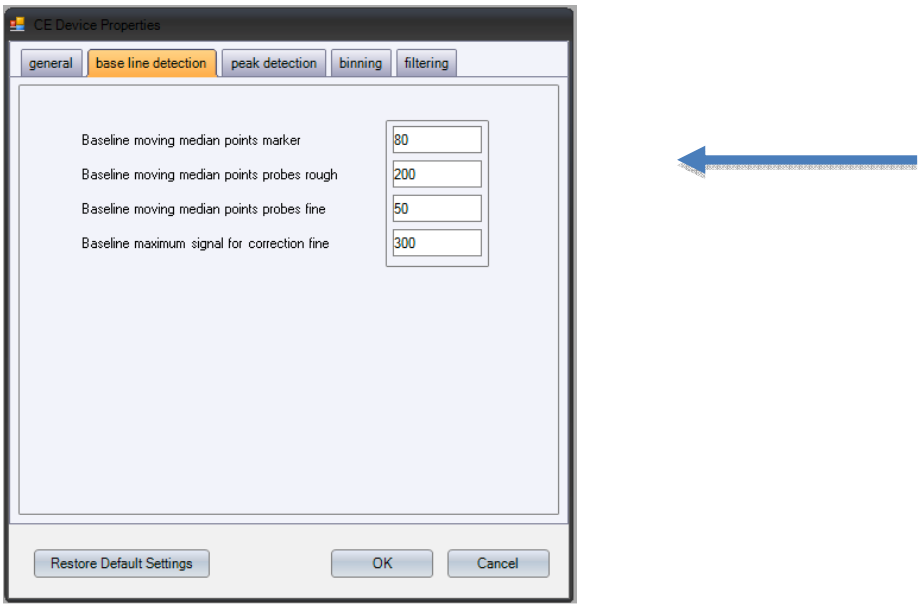


Figure 13 *Baseline settings.*

4.5.2 Peak detection settings

In capillary-based MLPA data analysis, peak detection is an essential step for subsequent analysis. Even though various peak detection algorithms for capillary electrophoresis data exist, most of them are designed for detection of peaks in sequencing profiles. While peak detection and peak size calling are very important processes for sequencing applications, peak quantification is not so important. Due to the relatively nature of the MLPA data, peak quantification is particularly important and has a large influence on the final results. Our peak detection algorithm exists of two separate steps; the first step exists of peak detection by comparison of the intensities of fluorescent units to set arbitrary thresholds and shape recognition, the second step exist of filtering of the generated peak list by relative comparison.

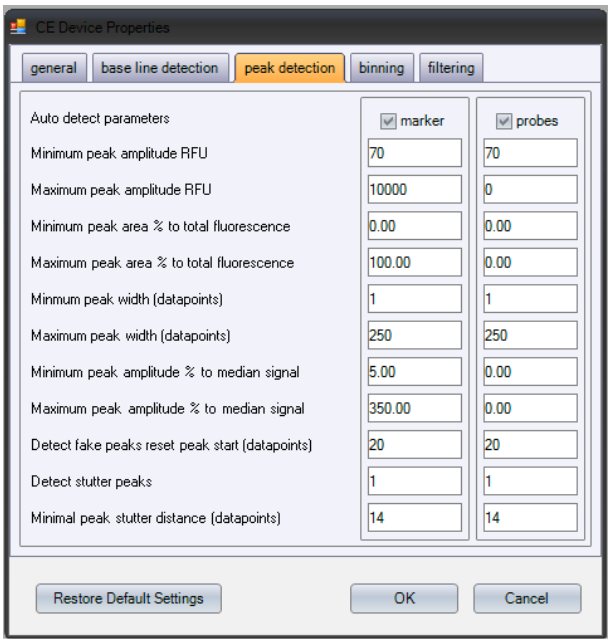


Figure 14 *Peak detection settings tab.*

Program administrators can modulate the peak detection algorithm thresholds for size marker channels and probe channels, by clicking on the second tab of the CE devices properties form (figure 14), which make use of the following criteria:

1) *Detection/Intensity threshold:*
This threshold is used to filter out small peaks in flat regions. The minimal and maximal peak amplitudes are arbitrary units and default values are provided for each different capillary system. These value are called the minimum and maximum peak amplitude RFU.
2) *Peak area ratio percentage*:
Peak area is computed as the area under the curve within the distance of a peak candidate. Peak area ratio percentage is computed as the peak area divided by the total amount of fluorescence times one hundred. The peak area ratio percentage of a peak must be larger than the minimum threshold and lower than the maximum set threshold. These values are called the minimum and maximum peak amplitude % to total fluorescence.
3) *Model-based criterion:*
The application of this criterion can consists of 3-4 steps:

• Locate the start point for each peak: a candidate peak is recognized as soon as the signal increases above zero fluorescence.
• Check if the candidate peak meets minimal requirements: the peak signal intensity is first expected to increase, if the top of the peak is reached and the candidate peak meets the set thresholds for peak intensity and peak area ratio percentage, then the peak is recognized as a true peak.
• Discarding peak candidates: if the median signal of the previous 20 data points is smaller then the current peak intensity or if the current peak intensity returns to zero. This value is called: “detect fake peaks reset peak start (datapoints)”.
• Detect the peak end: the signal is usually expected to drop back to zero designating the peak end. In some cases the signal does not return to zero, a peak end will therefore also be designated if the signal drops at least below half the intensity of the peak top and if the median signal of the 14 last data points is lower than the current signal. This value is called: “Minimal peak stutter distance (datapoints)”.
4) *Median signal peak filter:*
The median peak signal is calculated by the percentage of intensity of each peak as opposed to the median peak signal intensity of all detected peaks. Since the minimum and maximum thresholds are dependent on detected peaks, this filter will be applied after an initial peak detection procedure based on the criteria point 1-3. This value is called: “minimum and maximum peak amplitude % to median signal”.
5) *Peak width filter:*
After peak end points have been identified, the peak width is computed as the difference of right end point and left end point. The peak width should be within a given range. This filter is also applied after an initial peak detection procedure. This value is called: “minimum and maximum peak width (datapoints)”.
6) *Peak pattern recognition:*
This method is only applied for the size marker channel, and involves the calculation of the correlation between the data point of the peak top of the detected peak list (based on the criteria point 1-5) and the expected lengths of the set size marker. In case the correlation is less than 0.999, the previous thresholds will be automatically adapted and peak detected will be restarted. These adaptations mainly include adjustment of minimal and maximal threshold values. The minimal correlation quality is a default value and cannot be adjusted.

4.5.3 Binning settings

After each peak is detected it needs to obtain a size, which in general will be quantified in a number of nucleotides by a method called size calling. Size

calling is a method that compares the detected peaks of a MLPA sample channels against a selected size standard. Lengths of unknown (probe) peaks can then be predicted using a regression curve between the data points and the expected fragment lengths of the used size standard, resulting in a fragment profile.
Once all peaks have been size called, the profiles must be aligned to compare the fluorescence of the different targets across samples, an operation that is perhaps the single most difficult task in raw data analysis. Peaks corresponding to similar lengths of nucleotides may still be reported with slight differences or drifts due to secondary structures or bound dye compounds. These shifts in length make a direct numerical alignment based on the original probe lengths all but impossible. Our software uses an algorithm that automatically considers what the same peaks are between different samples, allowing easy peak to probe linkage. This procedure follows a window-based peak binning approach, whereby all peaks within a given window across different samples are considered to be the same peak. Our software algorithm follows four steps: reference profile analysis, applying and prediction of new probe lengths, reiteration of profile analysis and data filtering of all samples. The crucial task in data binning is to create a common probe length reference vector (or bin). While this procedure occurs completely automatically, some aspects may be adjusted in the CE devices properties window, under the binning tab (figure 15).

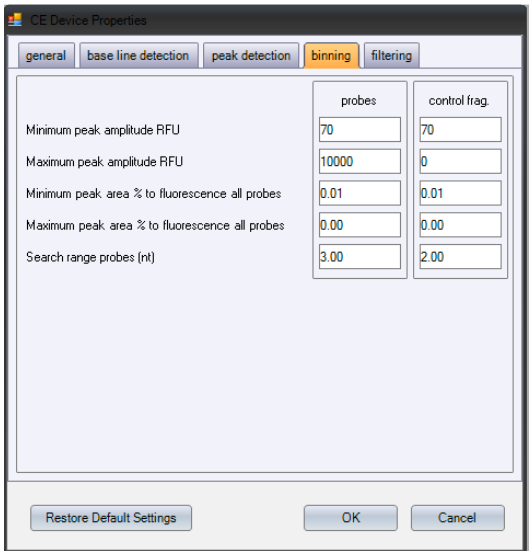


Figure 15 *Binning settings tab.*

Specific settings can be applied for control fragments and probe fragments such as:

1) *Detection/Intensity threshold*:
This threshold is used to filter out peaks which may be detected in the first step described at 4.5.1, but will compete in the automatic binning procedure. The minimal and maximal peak amplitudes are arbitrary units and default values are provided for each different capillary system. These values are called the minimum and maximum peak amplitude RFU at the binning tab.
2) *Peak area ratio percentage to all probe fluorescence:*
Peak area is computed as the area under the curve within the distance of a peak candidate. Peak area ratio percentage as compared to all probes is computed as the peak area divided by the total amount of fluorescence of all probes added times one hundred. The peak area ratio percentage of a peak must be larger than the minimum threshold and lower than the maximum set threshold to compete in the binning procedure. These values are called the minimum and maximum peak amplitude % to fluorescence all probes.
3) *Search range (nt):*
The search range determines the size in which the binning procedure will look for probes. If the minimal distance between probes is 6 nucleotides; the search range of each probe is 3 nucleotides plus/minus. The smaller this

search range the set, the more difficult it will be for the binning procedure to relate detected peaks with certainty to a probe.

4.5.4 Data filtering settings

Data filtering is the actual process where the detected fragments of each sample are linked with gene information to a probe target or control fragment. The binning procedure is thus only used to create common probe length reference vector, and not for filtering. The binning procedure may for instance only be applied on the sample that were set as reference samples, while the filtering procedure will be applied on all selected samples. Our algorithm assumes that peaks within each sample that fall within the same provided window or bin and have sufficient fluorescence intensity are the same probe. Our algorithm is also able to link more than one peak to a probe within one sample. The amount of fluorescence of each probe product may then be expresses the peak height, peak area of the main peak and the summarized peak area of all peaks in a bin. An algorithm can then be used to compare these metrics and decide which should optimally be used, alternatively users may set a default metric.
Filtering can be optimized by adjusting the settings shown in figure 16, these properties responds as equal to their values as the equally named properties do in the binning procedure, described in 4.5.3.

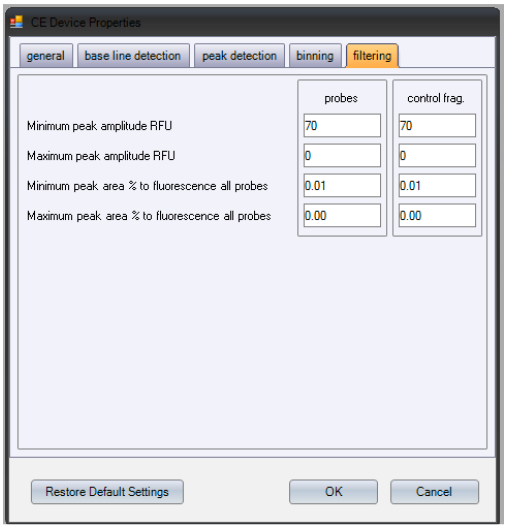


Figure 16 *Filtering settings tab.*

5. About projects

5.1 What does a project contain?

Our database setup contains a large number of subtraction levels, not only allowing users to efficiently store and review experimental sample data, but also allowing users to get integrative view on comprehensive data collections as well as supplying an integrated platform for comparative genomics and systems biology. While all data normalization occurs per experiment, experiments can be organized in a project, allowing advanced data-mining options enabling users to retrieve and review data in many different ways. Users can for instance review multiple MLPA sample runs from a single patient in a single report view. Multiple MLPA mix results may be clustered together, allowing users gain more confidence on any found results. The database can further handle an almost unlimited number of specimens for each patient, and each specimen can additionally handle an almost unlimited number of MLPA sample runs. By creating projects, users can furthermore collect data of different experiments in one project collection. In future updated version it will then be possible to create a summary of all data within one experiment, or for instance compare all data in a project against all data in another project.

5.2 Creating a new project

After creating an empty solution, users can add new or existing items to the empty solution by right clicking on the folder “Projects” in the organizations folder and selecting the option “add project” (figure 17).

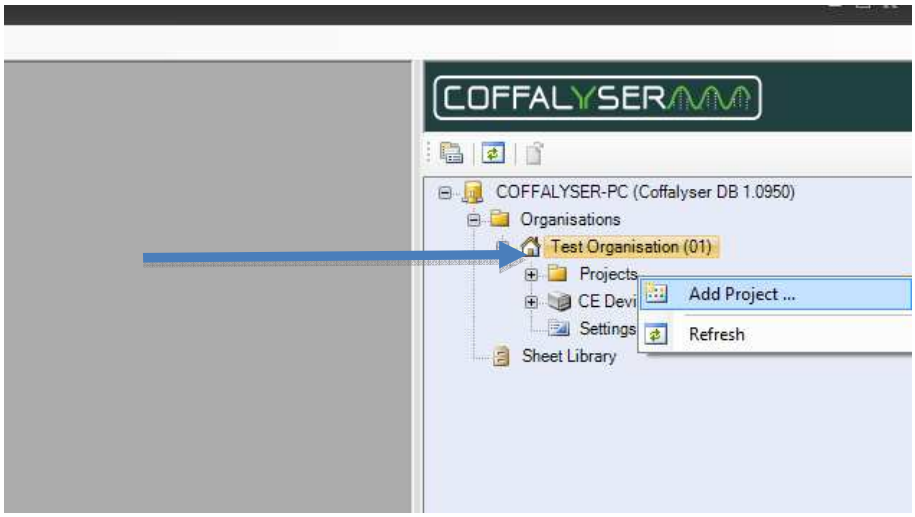


Figure 17 *Filtering settings tab.*

5.3 Project settings

After a new project is created you can define the default capillary electrophoresis device you wish to use and give a title to your project (figure 18). You can also fill in a short description about the project.

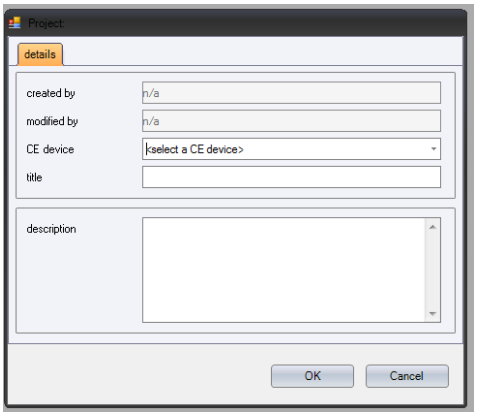


Figure 18 *Project settings window.*

6. About experiments

6.1 What does an experiment contain?

After creating an initial project, we can create experiment within this project. In each experiment data files can then be imported to the database and linked to this experiment. Users then need to define the experiment type and for each used channel or dye stream of each capillary (sample run) what the contents are. Each detectable dye channel can be set as a sample (MLPA kit) or a size marker. Samples may further be typed as: MLPA test sample, MLPA reference sample, MLPA positive control, or MLPA digested sample.

6.2 Creating a new experiment

To create a new experiment within a project, right click on the project you wish to add the new experiment to, and select “Add experiment” from the right click menu (figure 19).

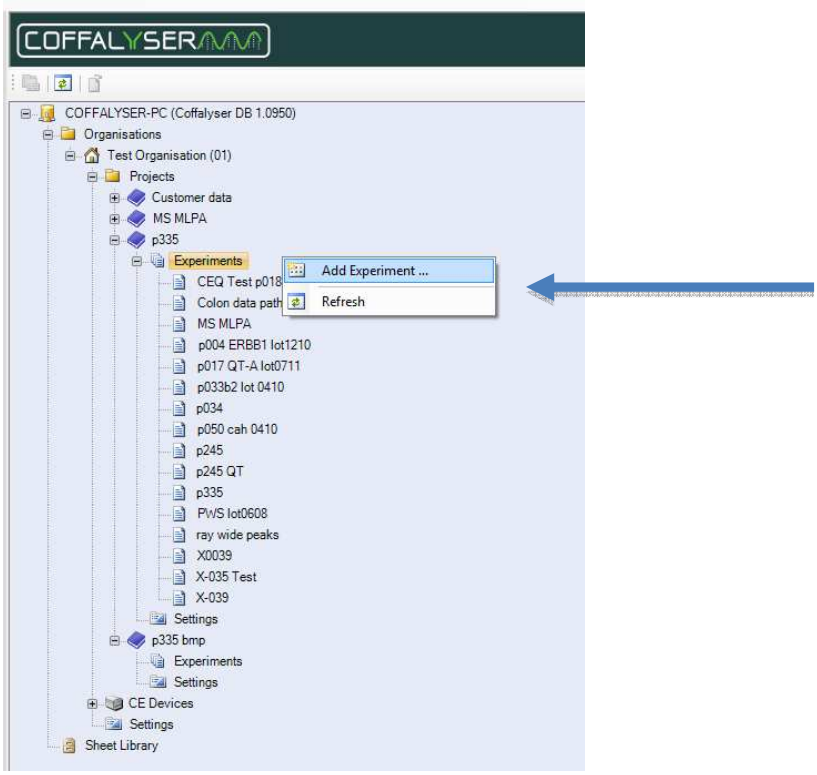


Figure 19 *Adding a new experiment in the database exploration window.*

6.3 Experiment settings

Directly after you create an experiment you will be able to adjust the experiment settings and give the newly created experiment a name and description (figure 20). The capillary electrophoresis device should already be filled in to be the default machine for that project. You may however choose to also include different machines within one project. After you click ok, the experiment will be created in the database, allowing you to continue to define the content of each channel.

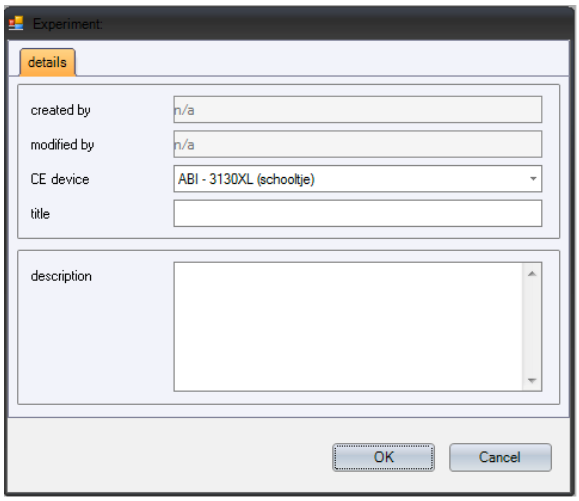


Figure 20 *Experiment settings window.*

6.4 Setting the experiment type

After adding a name and description to your experiment, you may define the settings required to start the fragment analysis in the next form (figure 20). First we need to determine what type of experiment we are analyzing. There are basically 3 types of experiments, these being:

1) Copy number analysis ("DNA/MLPA [default]"):
These are experiments that are performed using standard MLPA probes or custom probes that are designed according to the same rules. The used probes can only produce signals that are proportional to the amount of the DNA target sequences present in each sample. These experiments furthermore require data obtained from reference samples that were performed in the same experiment. This reference sample is usually performed on a sample that has a normal (diploid) DNA copy number for all target sequences.
2) Copy number / methylation status analysis ("DNA/MS-MLPA"):
These experiment are combined experiment where both the copy number and methylation status of the probe target sequences are calculated in a single analysis. While the copy number part is equal to that described at point 1, the methylation status analysis, requires a digested sample result together with each standard MLPA sample result. For MLPA probes that contain HHA1 sites the methylation status can then be determined by comparing the signal that is proportional to the amount of the DNA target

sequences present in each sample after digestion to the signal of the same target sequence of the same sample without digestion. In case only one of the two copies is methylated, the amount of target sequences available in the digested result will be 50% lower as compared to that of the undigested result.

3) RNA analysis ("RNA"):

RNA experiments are quite similar to copy number experiment, except that the probe target sequences are directed to RNA sequences. Sample DNA is therefore often purified from genomic DNA in order to minimize contamination and required reverse transcriptase. In the analysis you may also set reference samples (e.g. zero control, or RNA from control tissues) in order to make a relative comparison. Alternatively users may only evaluate the intra-normalized results, thereby comparing each probe signal against one or two reference probe signals within the same sample.

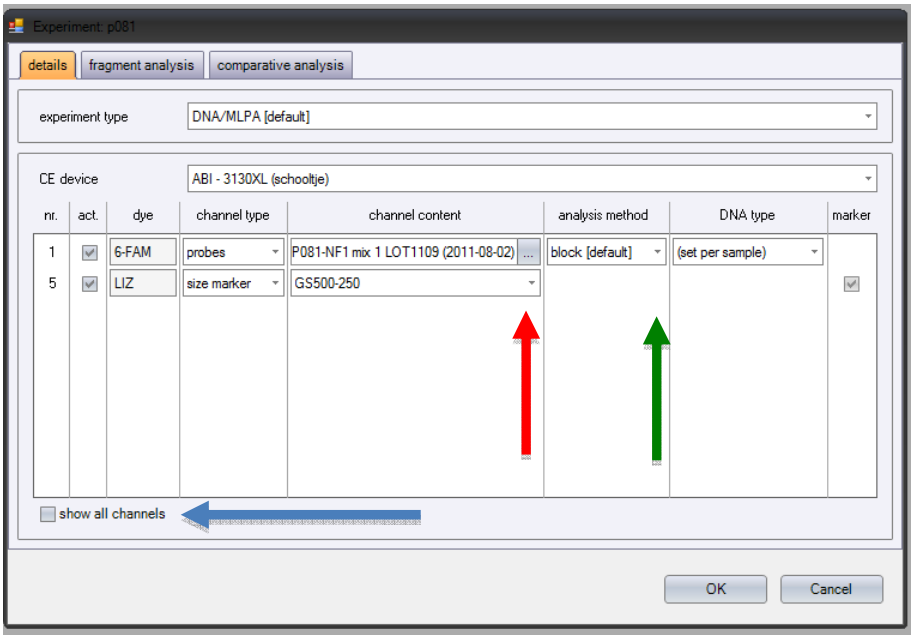


Figure 21 *Experiment fragment analysis settings window.*

6.5 Setting the channel contents

After settings the correct analysis method, you need to set for each dye channel what the expected contents are. The channels are usually set

correct as determined by your filter set. If your channels are not set correct, then click on the option box: “show all channels”, you will be able to select which channel you are using by ticking the option boxes in the first column called “nr” (blue arrow; figure 21). The name of each dye should appear in the next column. You will not be able to change the dye names since they are related directly to your filter set. Now you will be able to set the content type or “channel type” for each of your used channels by clicking on the dots or on the little arrow on the left side of the combo box in the channel content column (red arrow; figure 21). Channels are either set to “probes” indicating that in this channel peaks that can be related to a MLPA probe mix can be found; or the channel type can be set to “size marker”; indicating that this channel contains a size standard which can be used to compare the detected peaks against and give them a length in nucleotides (also see our FAQ). If you have set the content of a channel to be a probe mix, then you also need to define the products, lot and version number by using the probe mix selection form, which will appear after selecting the dots (red arrow; figure 21).

6.6 Setting the channel settings

After settings the channel contents you will find some other settings behind the channel type. In case you have indicated that you are using a “probes” channel type you also need to set an analysis method for the probe mix. The default method that will appear in most cases is “block [default]”. Block analysis means that the available reference probes are used to normalize the samples against the reference samples. Normalization in this case referrers to the division of multiple sets of data by a common variable in order to cancel out that variable's effect on the data. Reference probe are usually targeted to chromosomal regions that are assumed to remain normal (diploid) in DNA of applicable samples. In case a MLPA kit does not contain any reference probes, users may define their own reference set (see section 3.4 & 3.5) or use population method instead. In population analysis mode, all probes are used for normalization; this method is therefore only recommended in case the number of aberrations in each sample is expected to be very low (e.g. 1-2 aberrant probes target sequences in each sample). To change the analysis method click on the little arrow row define as a “probes” channel type (green arrow; figure 21).
The last two columns “DNA type” and “marker” will automatically be set for your and require no more adjusting. If you chance to work with more than 2 channels in one capillary sample runs, please see our advanced analysis section for more information about this.

7. About the fragment analysis

7.1 Importing the data files

After you have set the settings on the details page, you can go to the next tab where you can import your sample files. Right click anywhere on the screen and select "*Add (from file)"* from the right click menu (figure 22).

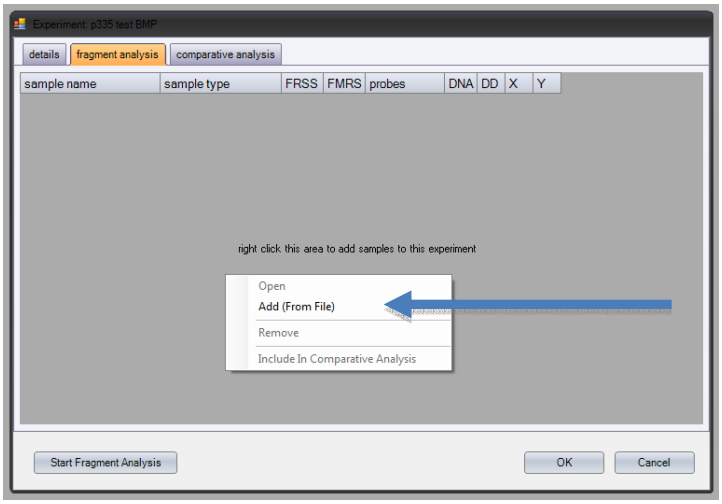


Figure 22 *Fragment analysis window for importing samples*

After selecting "*Add (from file)"* the file / folder import window will appear (figure 23). Here you can import files or complete folders into the database and automatically link them to the current experiment. For ABI-devices, ABIF files from all series can be imported (*.*fsa extensions); for CEQ-devices (Beckman) data from the CEQ-2000, CEQ8000 and CEQ8000 can be imported (*.*SCF or *.*esd extensions); for Megabace-devices data of all series can be imported (*.*rsd extensions) and for Agilent-devices data of the Bioanalyzer can be imported (*.*xml extensions). Select the "Add files" or "Add folder" (blue arrow, figure 23) and then select the files you wish to import in the explorer window. At this point the files are not stored in the database yet, click on "Import" (red arrow, figure 23) and to decode the

binary files and save them in the database. If all samples were imported correct, you can close this window to make the sample specific settings.

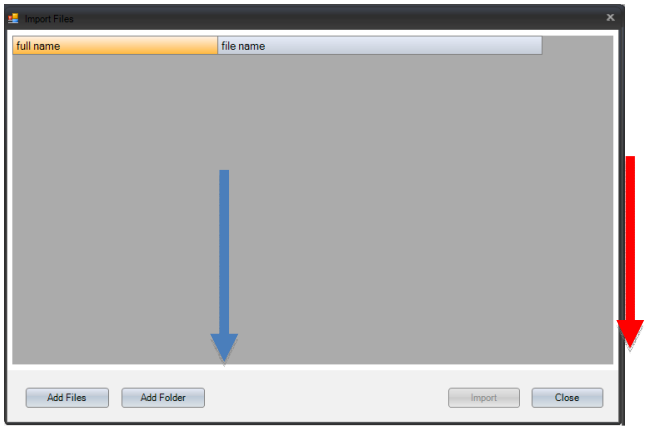


***Figure 23**. File / folder import form*

7.2 Starting the fragment analysis.

After importing your samples and you have closed the file / folder import form", the fragment analysis sample setup window will appear (figure 24). This form allows you to adjust the sample types that you have used in your experiment. You can set 4 different sample types; either by using the key-shortcuts or by changing the combo box by double clicking on the cells in the second column called "sample type". We distinguish the following types:

1) Samples or test samples ("key = s"), which will be normalized against the reference and are considered to be the unknown samples of which we want to know the copy number status of the test probes. For these samples we assume that the target sequences of the reference probes are normal or diploid for all autosomes or have an equal copy number as compared to the reference samples. In case no reference samples are defined in the experiment, each sample will be used as a reference. The data for each test probe of each sample will be compared to each other sample, producing as many dosage quotients as there are samples. The final ratio will then estimated by calculating the median over these dosage quotients.
2) Reference samples ("key = r") are used to display the balance of the measured signal intensities between sample and reference. The data for each test probe of each sample will be compared to each available

reference sample, producing as many dosage quotients as there are reference samples. The final ratio will then estimated by calculating the average over these dosage quotients. In case no reference samples are set, each sample will be used as reference and the median over the ratios be calculated. Next to this reference samples are used to estimate the effect of sample-to-sample variation on probe ratios of test probes by calculating the reproducibility of these probes in the reference sample population. These calculations may be more accurate under circumstances where reference samples are randomly distributed across the performed experiment.

3) Positive reference samples (“key = p”) are used to make an estimation of the behavior of a probe within a sample population with a known aberration. We can do this by calculating the distribution statistics for each probe over all sample ratio results of the same type. Next each unknown test sample result can be tested against several variables of that distribution, such as: the average, median, standard deviation, CV and 95% confidence range in order to calculate the probability that an unknown sample is equal of different to the distribution results of that sample type.
4) No DNA or blank controls (“key = b”) are analyzed MLPA experiments that do not contain any DNA. They are used to make sure not contamination has occurred during the performance of the experiment.
5) Digested sample are all samples that were digested (“key = d”) during the experiments and are used only to estimate the methylation status of each target sequence.

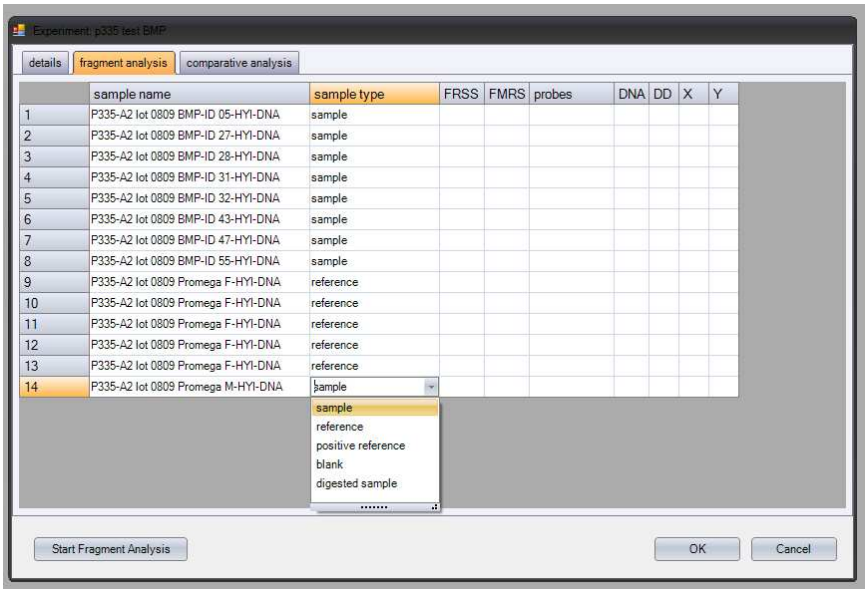


Figure 24 *Fragment analysis sample setup window.*

When you are finished adjusting all the sample types, click on the button called “Start fragment analysis” to perform the all-necessary steps to qualify and quantify each of the probe signals. The screen will automatically update and present the quality scores for each sample after the analysis is finished.

7.3 Fragment analysis settings

After you click on the fragment analysis button the fragment analysis settings screen will open (figure 25). This screen will allow you to change the main / advanced fragment analysis settings that are unrelated to the CE-device settings. The form consists of three pages related to different processes of the fragment analysis. First tab contains all settings related to the peak recognition, at the top you can set whether to use a basic baseline correction or an *advanced baseline detection method*. The exact differences and effect can be viewed in the fragments explorer at the fragment analysis steps tab as discussed later in this chapter. Enabling the advanced baseline method has to effect that the peak detection method is repeated twice. First baseline correction and peaks detection are performed as discusses earlier, following this the process is repeated, the second baseline correction however will now be made based on all the signals that are known to be unrelated to

peaks. When the baseline gets to an area known to be related to peak, the underlying baseline will be predicted towards the next point in the data streams that was unrelated to a probes. The *advanced baseline detection max degrees* determines how much the degrees increase the predicted line may have. In case the increase is more than the set percentage, the line will be predicted to the next possible point that has a less steep increase. This method ensures that the baseline will be corrected as close as possible, but that peaks that are so close that they are shaped into one peak are not split. Instead such peaks will be assumed to be split peaks and the fluorescence of these peaks will be divided over the two by splitting them in two on the lowest points in between the two peaks.
The peak recognition tab also allows you to change the way peaks are being size called. Coffalyser.NET allows 4 different regression types:

1. Linear regression: In statistics, linear regression is an approach to modeling the relationship between a scalar variable y and one or more explanatory variables denoted X. The case of one explanatory variable is called simple regression. Linear regression refers to a model in which the median, which the conditional mean of y given the value of X is an affine function of X.
2. Least squares Local median regression: LS local median regression refers to a model in which the median of the conditional distribution of y given X is expressed as a linear function of X. This makes the lines more robust against outlier than regular linear regression. By using the field *regression local size* the number of local points may be determines.
3. Polynomial regression: polynomial regression is a form of linear regression in which the relationship between the independent variable x and the dependent variable y is modeled as an nth order polynomial. Although polynomial regression fits a nonlinear model to the data, as a statistical estimation problem it is linear, in the sense that the regression function E(y|x) is linear in the unknown parameters that are estimated from the data. The order can be changed by changing the field *regression polynomial degree*.
4. Local linear regression: this methods closely resembles the LS local median regression but instead of feeding the model median values of conditional distribution of y, regression coefficient are determined locally by the actual data points. This results in a regression line with a differential coefficient at different points and is in effect not a straight line as the LS local median regression line is.

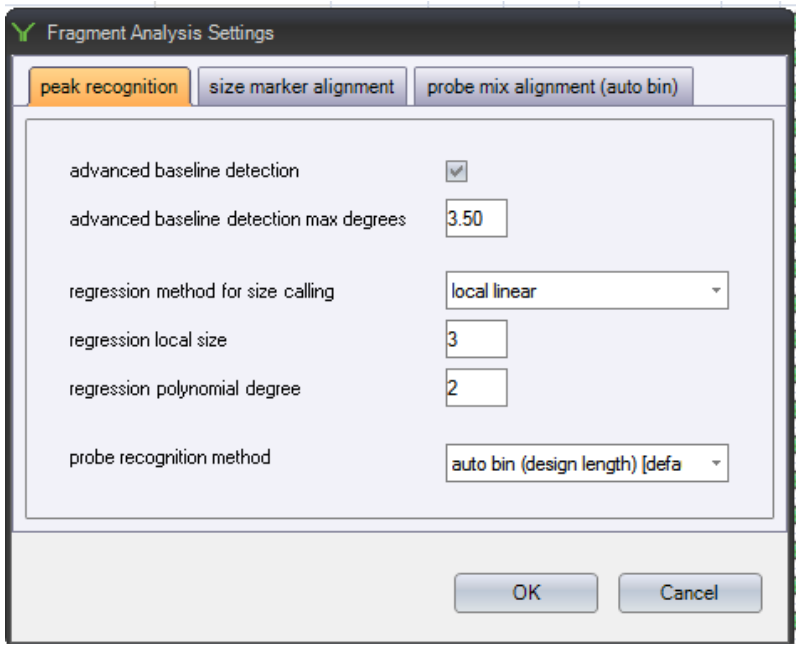


Figure 25 *Fragment analysis settings.*

As discussed earlier (but also in the coming chapter about manual binning), Coffalyser a window or panel based approach to link peaks with comparable lengths to the same probe. In order to define these panels or bins we need to compare the peak information we have, with what is expected. This is done automatically with an auto bin procedure. The more the lengths are comparable to the found lengths, the more chance that the procedure will find all probes successfully. Coffalyser allows two types of probe length to be used for the auto bin procedure: the probe design length, which are the real length of the fragments and the Coffalyser lengths. Coffalyser lengths are lengths that are filled by MRC-Holland to make the binning procedure more successful since they are based on the detected lengths found during the quality tests. Finally you can also filter you data based on a manual bin set by selecting the manual option. How to create a manual bin set is explained in the section "*creating a manual bin set for data filtering*".

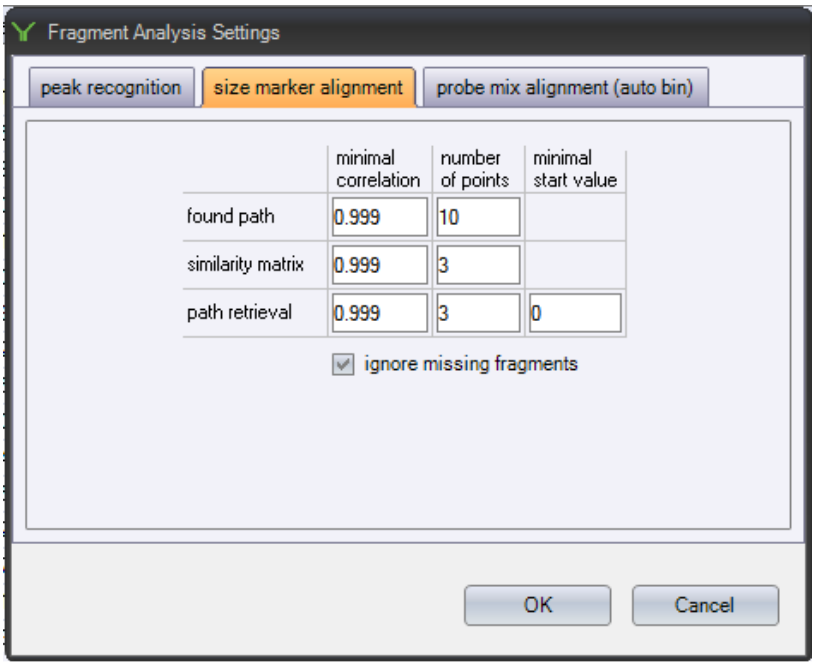


Figure 26 *Fragment analysis size marker alignment settings*

On the second tab of the fragment analysis settings form you can find the size marker alignment matrix settings. These settings are in general optimized for the user and do not require adaption. This page describes the properties that are used to auto detect which detected peaks in the size marker stream can be related to the expected fragments. To do this Coffalyser uses several matrix correlations calculations. From the top down we first find the minimal correlation and number of local points that should be used for that correlation. This is the minimal correlation between the data points of the found peaks with the lengths of the expected fragments using a local correlation estimator (in other words, we do not expect the complete line to be linear with a high correlation, but each 10 consecutive points. The similarity matrix refers to the method of peak-probe selection, this is done by creating a matrix of all peak data points versus all expected fragment lengths. Then all local correlations of a number of local points (3 in figure 26) with the crossed lengths is calculated diagonally. In case a correlation was found over position in the matrix, each correlation that is higher than the minimal demand scores 1. In case some position have already a score of one, and an overlapping correlation is found, the score will be plus one. This

is a supporting matrix for the path retrieval method. After this similarity matrix is made, path retrieval is performed by starting at the corner with the largest data point and length. From this point a path will be tried to find follow a number of rules. Basically these rules involve the movement of a peak in the size marker stream and relating it to the expected length of a size marker peak. Going through the matrix you may then find the next peak at a position -1 /-1 of the previous peak, but we will only move that way in case the correlation meets the set requirement of path retrieval together with the number of points. Using this method we may skip both background peaks but also ignore expected size marker fragments if required. After a complete path is found back, we measure the correlation between the peak data points and the expected size marker lengths to make sure that the quality is ok. This result will then be stored in a new matrix and the path retrieval method is repeat for all position in the similarity matrix, or for all positions in the similarity matrix for which a minimal start value was found, as set by the path retrieval minimal start value. Finally the path that has the same length as the size marker and for which a high correlation was found for the found path, will be used for the size calling procedure.

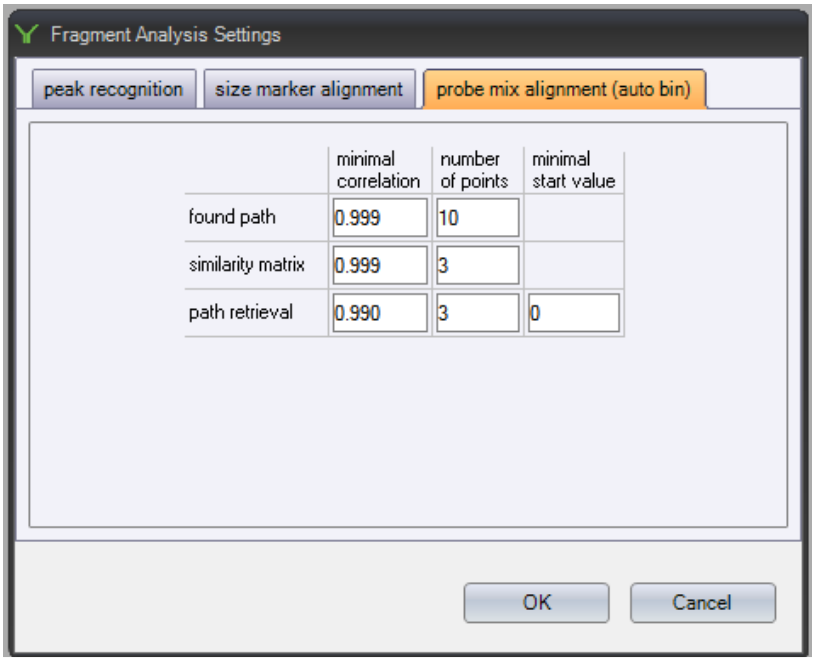


Figure 27 *Fragment analysis probe alignment (auto bin) settings*

On the final tab of the fragment analysis settings form you can find the probe mix alignment (auto bin) settings. These settings are in general optimized for the user and do not require adaption. This page describes the properties that are used to auto detect which detected peaks in the probe streams can be related to the expected fragments. The method is essentially equal to the earlier described method for recognition of the size marker, we correlate data points of all detected peaks in the probe channels against the probe design length or coffalyser lengths. Note that the path retrieval correlation is less high than the same settings for the size marker. This is done because the migration patterns of MLPA probes are not yet optimized and the mobility to length correlation tends to deviate a little. This will be optimized in the future by completing all Coffalyser length and then the correlation may also be increased.

7.4 About the fragment analysis quality scores.

Because of problems arising from poor sample preparations, presence of PCR artifacts, irregular stutter bands, and incomplete fragment separations, a typical MLPA project requires manual examination of almost all sample data. Our software was designed to eliminate this bottleneck by substantially minimizing the need to review data. By creating a series of quality scores to the different processes users can easily pinpoint the basis for the failed analysis. These scores include quality assessment related to: the sample DNA, MLPA reaction, capillary separation and normalization steps (figure 28). Each collective quality score, or score that summarizes a number of aspects or factors starts with 100 points which can be correlated with high quality (or green). Depending on the importance and found severity of abnormality of each factor a number of penalty points are being given for each measured quality factor. The quality of each step can fall roughly into three categories.

1) High-quality or green. The results of these analysis steps can be accepted without reviewing.
2) Low-quality or red. These steps represent samples with contamination and other failures, which render the resulted data unsuitable to continue with. This data can quickly be rejected without reviewing; recommendations can be reviewed in Coffalyser.NET and used for troubleshooting.
3) Intermediate-quality or yellow. The results of these steps fall between high- and low- quality. The related data and additional recommendations can be reviewed in Coffalyser.NET and used to optimize the obtained results.

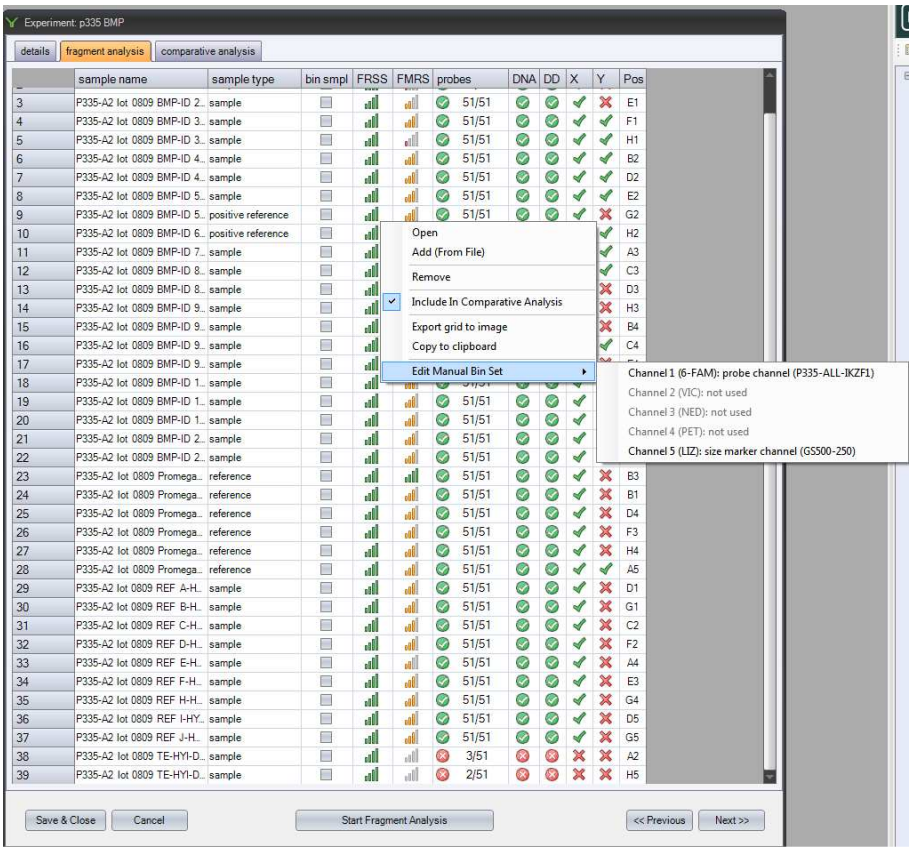


***Figure 28**. Fragment analysis quality scores and right click menu.*

Based on the quality scores you may use the right click menu to open the fragment analysis results explorer, add or remove samples and include sample for the comparative analysis.

7.3.1 FRSS

FRSS: Fragment Run Separation Score displays the quality of the fragment separation and peak sizing quality by evaluating the quality of the peaks in the size marker channel. To get to a final score several different criteria are evaluated that each have a penalty weight, which is subtracted from 100 start points or 100% ok. Each score that is dependent on the measurement of signal intensities has adjusted criteria that are dependent on the machine type. The method of quality assessment may thus different between

machines, to find the exact criteria for each machine for the different quality control checks please check the tables in the appendix.

7.3.1.1 FRSS check 1: Correlation of the size marker curve

Background: the techniques described in this section are used to investigate relationships between two variables (x and y). Is a change in one of these variables associated with a change in the other? For MLPA the size call correlation refers to the correlation between the relative migrations of the size marker fragments, in data point or time, with the expected fragment lengths of the used size marker in nucleotides. We can use the technique of correlation to test the statistical significance of the association. In other cases we use regression analysis to describe the relationship precisely by means of an equation that has predictive value. We deal separately with these two types of analysis - correlation and regression - because they have different roles.
The Correlation can be calculated by:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}},$$

Problems are indicated when: without a good correlation, no proper length estimation of unknown peak fragments can be performed. We therefore require a correlation of minimal *0.999* to continue with size calling. If case the correlation is lower than *0.999,* size calling may still be successful but probe lengths may deviate more than 0.5 nt from their corresponding partners in other capillaries. For correlations lower than *0.999*, a subtraction will therefore be executed of 80 points, resulting in a direct fail of this run.
Recommendation with problems: check the detected size marker peaks pattern and the peak detection settings for the marker. If peaks are not present or similar peaks exist in the pattern that are undistinguishable from the original size marker peaks, samples need to be rerun. Problems with peaks that are detected and should not have been detected or peaks that are not detected but are present in the raw data can usually be resolved by adjusting the peak detection settings of the marker.

7.3.1.2 FRSS check 2: Baseline of the size marker channel

Background: high baseline levels can lead to erroneous base calling and short read lengths. High baselines furthermore decrease the dynamic range of detection of that channel. Optimal performance of the capillary system is achieved when the baselines for all channels are below 5% of the maximum detectable intensity.
Problems are indicated when: in case the measured average baseline, or signal intensity of the size marker specific dye stream without running fluorescent products, is above 10% of the maximum intensity of the machine a warning will be given and 15 points subtraction on the FRSS total. In case the baseline is between the 7 and 10% only a notification will be given and 10 points will be subtracted from the FRSS total. For an ABI-3130xl for example: the maximum baseline intensity for the marker is set at 700 units for a warning and 560 units for a notification.
Recommendation with problems: Remove the capillary array at the manifold end and clean the capillary window. Use sterile water. DO NOT USE METHANOL. Clean in one direction only. Use “Direct Control” from the “Run” application to purge the manifold and to fill the capillaries with new gel and then clean the capillaries.

7.3.1.1 FRSS check 3: signal Intensity of the marker peaks:

Background: while peak detection and peak size calling are very important processes for sequencing applications, peak quantification is not so important. Due to the relatively nature of the MLPA data, peak quantification is particularly important and has a large influence on the final results.
Most capillary electrophoresis devices use electro kinetic injection procedures to introduce the sample into the flowing mobile phase which differ from LC in two ways: the injection volume is not as well defined and the injection is performed with the electric field turned off. Both of these features can contribute to quantitative errors of analysis. Because the entire internal volume of a 50cmx50 um-inner diameter capillary is only 981 nL, the injection volume must be kept quite small. Larger volumes will have more band broadening and band broadening may further be effected caused by diffusion. Lower strength ionic solution as the sample diluents may allow sample stacking and permits larger volumes for injection.
Since we are not interested in quantification of the size marker peaks, but only use the size marker for comparing the relative migration of the size marker fragments with those of the probe fragment in the same lane, the amount of the size marker should be as minimal as possible allowing more injection and better quantification of the MLPA fragments. Signal strength is important to be properly visible, because without sufficient signal, it is very unlikely that accurate base calls can be made. Optimal size marker signals

fall between the 1-10% of the detectable maximum and should be at least 3x the signal of the baseline.
Problems are indicated when: in case the measured median peak signal intensity of the recognized size marker peaks is below 1% of the machine detectable maximum or above 70% of the absolute maximum a warning will be given and 20 points will be subtracted from the FRSS. Notifications will be given in case the median peak signal intensity is between the 1 and 1.25% or between the 60 and 70% of the machines absolute maximum. For an ABI-3130xl for example the minimum intensity for the marker median peak signal intensity is 100 units for warning and 125 units for a notification, the maximum median peak intensity is set at 5600 for a warning and 4800 for a notification. It should be noted that the minimum demands are often set by the maximum minimum intensity for proper peak quantification and not so much as the percentage of the detectable maximum.
Next to the median peak signal intensity we also check the maximum peak signal intensity of all detected marker peaks. In case the measured peak maximum intensity of the recognized size marker peaks is above 87.5% of the absolute maximum a warning will be given and 15 points will be subtracted from the FRSS. Notifications will be given in case the maximum peak signal intensity is between the 70 and 87.5% of the machines absolute maximum and 10 points will be subtracted from the FRSS total. Fr an ABI-3130xl the maximum peak intensity is set at 5600 for a notification and at 7000 units for a warning.
Recommendation with problems: adjust the concentration of the size marker in the injection mixture to increase or decrease the signal intensities of the marker

7.3.1.4 FRSS Check 4: Signal drop of the internal run of the size marker fragments

Background: size markers are usually developed having fragments concentrations in equal amounts of all peaks, the reproducibility of the separation method may thus be examined by evaluating the intensity of the size marker fragments. In addition, the presence of the same multiple bands in several lanes in different regions of the gel provides information regarding possible lane-to-lane variation in the electrophoresis migration of sample material. Most markers (gs-500, 600-CEQ) are designed to give equal signal intensities over all fragments. A drop in signal of the fragments is thus probably introduced by the capillary electrophoresis and will also have a similar effect on the MLPA probes.
Problems are indicated when: To make sure that the signal drop is caused by a problem during the separation we combine a check on the drop of signals intensity over the run together with the measurement of the widening

of the peak signals. In case there is a problem with the capillaries in most cases a signal drop is always accompanied by widening of the peaks. First we measure the percentage of signal drop by comparing the median signal intensity of the first half of all size marker peaks to the median signal intensity of the last half of all size marker peaks. Then we measure the amount of peak widening by taking the first quartile of all measured peak widths and comparing this with the widest peak in that run. In case signal drop more than 60% a warning will be given and 30 point will be subtracted from the FRSS, in case this percentage is 40, a notification will be given and 15 points will be subtracted. In case this sloping is accompanied by peak widening of more than 50% another 35 penalty points will be given to the FRSS total.
Recommendation with problems: rerun with alternative injection settings. Run also a lane with only marker, since contaminant introduced from the sample DNA may also have an effect.

7.3.1.5 FRSS Check 5: Size marker complete / incomplete

Background: we usually expect that all fragments of the used size marker will be visible and detectable. In some cases however the marker may only be partially present or there is too much noise surrounding certain fragments to properly recognize the size marker peak.
Problems are indicated when: not all marker peaks that were expected were found but size calling was performed because the correlation of the remaining size marker peaks length with their data points still had a good correlation. Even though and incomplete marker with a good correlation does not necessarily have to cause problems, it does require good manual examination of the data. Runs with an incomplete marker therefore get a subtraction of 60 points of the FRSS total.
Recommendation with problems: rerun with alternative injection settings. Run also a lane with only marker, since contaminant introduced from the sample DNA may also have an effect.

7.3.2 FMRS

FMRS: fragment MLPA reaction score displays the quality of the performed MLPA reaction. To get to a final score seven different criteria are evaluated from the probe mix channel. Start score of the FMRS is 100 points.

7.3.2.1 FMRS check 1: signal Intensity of the probe fragments

Background: too little template leads to poor signal strength, which in turn leads to poor base calling and increased variation and thus unreliable results. Too much DNA results in a greater number of short extension fragments during labeling, which are preferentially loaded into the capillary during sequencing. This will result in a signal to size drop. Excess DNA may also result in lower signal strength since it will compete with the labeled DNA for injection leading to poor resolution again leading to unreliable results. Overloading, and more specifically truncated peaks will result in a complete fail of the quantification of the fragment, since only part of the product will be measured. Capillary system usually use CCD camera's to measure the amount of fluorescence so over or under loading of sample can be a problem for signal quantification. The optimal range for peak quantification is quite limited as compared to the dynamic range of most devices and it is thus crucial that most MLPA probe signal are in the optimal range when using it for copy number assessment.
Problems are indicated when: for most devices we give warning in case the median probe signal is below 4% of the maximum intensity or above 60% of the maximum signal intensity, resulting in a subtraction of 20 points from the FMRS. In case the signals are between the 4-5% or 50-60% of the absolute maximum the penalty is only 10 points. In case of an ABI-3130XL the minimum signal intensity of the median probe signal is 300 units and the maximum intensity is 5000 units.
Recommendation with problems: low raw data signal can be caused by a variety of issues. One of the most common causes is lack of sufficient DNA template in the cycle reaction. It is vital to the success of fragment analysis to have the correct amount of template in the reaction. It is recommended that 50–100 fmoles of product DNA be used in the cycle sequencing reaction. This provides enough template to generate an adequate amount of fluorescently labeled DNA sequencing fragments yet not so much as to cause current problems. Half this amount of DNA template (25–50 fmoles) should be used for single stranded DNA templates such as M13 phage DNA and even less DNA is needed for small PCR products (**10–50 fmoles** for PCR products less than 3KB in length). In many cases the amount of template added to the reaction is not determined and therefore, insufficient template is present. In other cases, an incorrect approximation of the DNA concentration is made. Spectrophotometer estimation of DNA samples is only valid if the DNA is pure (as in the case of commercial DNA template purification methods). Crude preparation of DNA templates which have substantial amounts of protein and/or RNA will over estimate the concentration of the template and cause the user to add too little DNA to the MLPA reaction (as in the case of crude alkaline lysis minipreps). Corrective actions: Add the correct amount of the template DNA to the reaction. This will require quantification of the template DNA by spectrophotometer (in the

case of commercial DNA preparations) or by estimation using agarose gel electrophoresis and comparison to a known quantity of DNA. Alternatively, the user could try a dilution series with the same template starting with an amount that is obviously too high and ending with an amount which is much too low. This method assumes that the user knows the approximate amount of template added to the reaction (this may be from previous work using similar DNA preparation methods). Use the preheat treatment for highly super coiled DNA. Most commercial DNA preparations yield highly super coiled DNA. The preheat treatment will knick the super coiled DNA which yields much more efficient DNA reactions (linear molecules sequence better than super coiled molecules).

Low raw data signal due to “bad formamide”; formamide is used to resuspend the DNA sequencing fragment prior to loading on to the electrophoresis deice. The formamide solution must be prepared and stored properly to achieve high quality sequencing data. If the formamide is not deionized and stored properly it will decompose into ammonia and formic acid. The formic acid then destroys the fluorescent dyes and produces low Raw Data signal Corrective actions: Use the special Sample Loading Solution (SLS). Do not freeze-thaw the SLS or formamide solutions. Store aliquots at –20℃ in anon-frost free freezer and use the aliquots only once.We do not recommend using water to resuspend the DNA sequencing fragments prior to loading. Some dyes are not stable inpure water solutions and will yield Raw Data signals similar to that of “BadFormamide”.

Low Raw Data Signal Due to Insufficient Sample Injection; poor injection of DNA fragments onto the CEQ capillaries will lead to low Raw Data signal. Since the CEQ uses electrokinetic injection it is highly sensitive to excess salts in the loading solution. The excess salts compete with the DNA sequencing fragments during injection and result in lower loading of the fluorescently labeled DNA molecules. The sources of the excess salts are improperly purified sequencing reactions and decomposed formamide. Corrective actions: Follow a desalting procedure such as: ethanol precipitation. If using spin column purification methods make sure that the column materials do not contain salts (check with the spin column manufacturer for details for using their products with capillary sequencers).

DNA Polymerase inhibitor; if the correct amount of template was added and the preheat treatment does not yield a substantial increase in Raw Data signal increase the number of cycles inthe thermal cycling program from 30 to 40 or 50. If the correct amount of template was added and the Preheat Treatment and / or increasing cycle number does not yield a substantial increase in Raw Data signal, a DNA Polymerase inhibitor may be present (do not resuspend DNA in DEPC treated water). In this situation further purification of the DNA template may be required. In some cases a simple ethanol precipitation of the plasmid will remove the inhibitor, whereas other

situations may require the use of commercial DNA preparation methods such as the Qiagen QiaQuick kit.
Low Raw Data Signal Caused by Poor Quality Mineral Oil; the mineral oil supplied in the DTCS kit is high quality oil containing no detectable nuclease activity. The use of other lower quality mineral oils can lead to sample degradation and hence low signal as shown below. The red and black dyes are particularly susceptible to this problem.
Recent experiments at MRC-Holland have shown that the use of patient or reference DNA samples with insufficient buffering capacity can result in abnormal MLPA results. These experiments indicate that a minimum of 5 mM (preferably 10 mM) Tris-HCl with a pH between 8.0 and 9.0 should be present in the 5 µl DNA sample before heating to 98 oC for DNA denaturation. Depurination of DNA at low pH and elevated temperatures is well known (e.g. PMID16412692; PMID10454625; http://openwetware.org/wiki/DNA_stability). This depurination is more severe at low ionic strength. DNA samples eluted from a purification column with water are therefore especially vulnerable. This depurination of sample DNA can have two effects:

1. No ligation of the MLPA probe oligonucleotides can occur when the sample DNA is depurinated at the ligation site of the MLPA probe, resulting in a lower probe signal.
2. Depurination of the sample DNA in the rest of the sequence detected by the probe can result in destabilization of the binding of the probe oligonucleotide to the sample DNA, resulting in a lower signal.

7.3.2.2 FMRS check 2: maximum probe signal Intensity of the sample.

Background: high probe signal intensity; in a few cases, the signal strength can be so high that it saturates the detector. This can lead to an erroneous base call where the software will artificially estimate peak height and position. In this case, the software inserts extra bases into the base sequence. By setting the raw data to full scale (CEQ = 137,000 counts) and looking at the peak shapes the user can determine if peaks are “over-ranged”. If the peaks are “squared-off” at the top, then the detector is saturated and the peaks are “over-ranged”.
Problems are indicated when: next to the median probe signal intensity we also check the signal intensity of the highest probe signal. A probe signal that is completely off scale may give a wrong ratio as compared to the reference and should thus be treated with caution. In case the highest probe signal surpasses 95% a warning is given and 15 points are subtracted from

the FMR|S. In case this signal is between the 90-95% only a notification is given and 10 points are subtracted from the FMRS.

Recommendation with problems: If the peaks are too high, the simplest solution is to rerun the same sample using a shorter injection time (for example: 7.5 seconds instead of 15 seconds). Use less template DNA or less thermal cycles to decrease the amount of fluorescence signal generated by the sequencing reaction.

7.3.2.3 FMRS check 3: Baseline Intensity of the probe dye.

Background: high baselines decrease the dynamic range of detection of that channel. Optimal performance of the capillary system is achieved when the baselines for all channels are below 5% of the maximum detectable intensity. In most cases we expect that problems with baselines are also visible in the size marker, however it may also be apparent that background fluorescence may be caused by the injected fluids itself and it is therefore impossible to resolve this issue with methods other that data analysis corrections.
Problems are indicated when: in case the measured average baseline, or signal intensity of a probe specific dye stream without running fluorescent products, is above 10% of the maximum intensity of the machine a warning will be given and 15 points subtraction on the FMRS total. In case the baseline is between the 7 and 10% only a notification will be given and 10 points will be subtracted from the FMRS total. For an ABI-3130xl for example: the maximum baseline intensity for the marker is set at 700 units for a warning and 560 units for a notification.
Recommendation with problems: Remove the capillary array at the manifold end and clean the capillary window. Use sterile water. DO NOT USE METHANOL. Clean in one direction only. Use “Direct Control” from the “Run” application to purge the manifold and to fill the capillaries with new gel and then clean the capillaries.

7.3.2.4 FMRS Check 4: Signal drop of the internal run of the probe fragments

Background: An effect that is commonly seen with MLPA data is a drop of signal intensity that is proportional with the length of the MLPA product fragments. This signal to size drop is caused by a decreasing efficiency of amplification of the larger MLPA probes and may be intensified by sample contaminants or evaporation during the hybridization reaction. Chemical remnants from the DNA extraction procedure and other treatments sample tissue was subjected to, may allot to impurities that influence the Taq DNA

polymerase fidelity. Alternatively target DNA sequences may have been modified by external factors, e.g. by aggressive chemical reactants and/or UV irradiation which may result in differences in amplification rate or extensive secondary structures of the template DNA that may prevent access to region of the target DNA by the polymerase enzyme (Elizatbeth van Pelt-Verkuil, 2008). Signal to size drop may further be influenced by injection bias of the capillary system and diffusion of the MLPA products within the capillaries. Even though some signal to size drop is expected, extreme drops may give problems due to the low signal intensity of the largest probes, furthermore if there exists a difference in size to signal sloping between the samples and references the ratio results will also be affected.
Problems are indicated when: to check if there are problems with probe signal to size drops we use a similar procedure as applied for the signal drop of the size marker. Measurements are again combined with a check on the widening of the probe signals. In case there is a problem with the capillaries in most cases a signal drop is always =accompanied by widening of the peaks. First we measure the percentage of probe signal drop by comparing the median signal intensity of the first half of all probe peaks to the median signal intensity of the last half of all probe peaks. Then we measure the amount of peak widening by taking the median of the first half of all probes peak widths and comparing this with the widest peak in that run.
In case of a signal drop of more than 70% a severe warning will be given and 60 point will be subtracted from the FRSS, in case this percentage is between the 60-70% a warning will be given and 35 point will be subtracted from the FRSS, in case this percentage is between 40-60% a notification will be given and 15 points will be subtracted. In case signal to size sloping is accompanied by peak widening of more than 50% another penalty will be given to the FRSS total. In case of a severe warning another 30 penalty points will be given, in case of a warning, 20 extra penalty points will be given and in case of a notification in combination with peak widening 10 extra penalty points are subtracted from the FMRS.
Recommendation with problems: rerun with alternative injection settings. If the size to signal drop is clearly not linear and not visible in the size marker, redoing the MLPA reaction with a lower sample volume or cleaned-up sample may provide better results.

7.3.2.5 FMRS Check 5: Percentage of unused primer

Background: in a successful MLPA reaction more than 70% of the added primer should be incorporated in probes. Often a lot of the available primer is caught away before the start of the MLPA reaction either by contaminants or DNA fragments that have complimentary sequences to one or both used

primers. If more than 30% of the primer is unused it may cause a drop in signal and thus can give unreliable results. A large primer-complex peak should be seen in the shorter length region of the profile.
Problems are indicated when: to measure if there is too much primer left we compare the fluorescence of the primer to the total amount of fluorescence of the probe peaks. In case a MLPA mix is expected to have 40 probe signals more than 40% primer will result in a warning and 40 points will be subtracted. In case this percentage is between the 20-40, a notification will be given and 15 points will be subtracted. Smaller mixes are allowed to have larger primer percentages; for mixes with 15-30 probes primer percentage criteria are increase with 10% and for mixes with less than 15 probes percentages may be 20% higher.
Recommendation with problems: either use different primers, or make sure that the PCR is started with a hot-start. MRC-Holland decided to use special primer blockers in combination with the PCR primers which circumvent the need for a hot start.

7.3.2.6 FMRS Check 6: Probes to peaks noise percentage

Background: the percentage of peaks that were detected, that were not recognized as MLPA fragments or probes is considered as noise. Large amount of background peaks may disturb the quantification of fluorescence of other probe related peaks. Large amounts of shoulder peaks may furthermore be caused by too large DNA concentrations or too high polymerase concentrations.
Problems are indicated when: if more than 70% peak signals are detected that were not recognized as probe signals a warning will be given and 20 points will be subtracted from the FMRS. In case the percentage of noise peaks is between the 40-70% a notification will be given and 10 points will be subtracted from the FMRS.
Recommendation with problems: increase the minimal peak detection thresholds. If the peaks are clearly visible and may cause problems with the quantification of other probe related peaks, then the product separation should be repeated or the MLPA reaction.

7.3.2.7 FMRS Check 7: Baseline curvature

Background: next to normal baseline heightening, baseline curving may also occur. Baseline curving is a heightening baseline at a local spot, most of the times directly below the probe signals. Most of this signal originates from the probe products but are often not proportionally to the rest of the run. Our baseline correction method resolved this issue by cutting the baseline through this curve which resolves the issue in most cases. Baseline

curvature however may still influence the peak height and area of all probe signals in that area. By measuring the fluorescence underneath the probes by a Zero-Baseline, which can be created by plotting a regression line through the data points in the beginning and end of each run, and comparing this to the fluorescence underneath the probes with our actual baseline we may find how much curvature there exists. This thus also indicated how much fluorescence we removed by not applying a straight baseline.
Problems are indicated when: if more than 50% of baseline curvature exists a warning is given and 40 penalty points are subtracted, in case baseline curvature is between the 30-50% only a notification is given and 15 points are subtracted.
Recommendation with problems: in most cases baseline curvature is likely caused by a high concentration of part of the injection products which are of similar size. This can be caused by inadequate mixing of the injection mixture or an injection bias e.g. by a too high injection voltage.

7.3.2.8 FMRS Check 8: DNA concentration check

Background: If the DNA concentration during the MLPA hybridization was insufficient for a reliable MLPA reaction unreliable results may be produced. MLPA reactions can be performed in a concentration range between the 20-500ng. We assume the DNA concentration is about 10ng if the median signal intensity of the Q-fragments is higher than the signal intensity of the 92 fragment / 3. We furthermore assume the DNA concentration was about 5 ng if the median signal intensity of the Q-fragments is higher than the signal intensity of the 92 fragment / 2. In case
Problems are indicated when: in case the ratio of 92 ligation fragment as opposed to the median signal of the Q-fragments is lower than 2; the DNA concentration was assumed to be too low and a warning will be given. Even though the DNA concentration is evaluated separately it will also affect the FMRS, warnings will minimize the FMRS with 60 points. A warning will be given if the ratio of the 92 ligation fragment as opposed to the median signal of the Q-fragments is between the 2-3.
Recommendation with problems: in case there is a clear problem with the DNA concentration, reaction should be repeated using higher DNA concentrations. If no higher concentrations are available samples may be concentrated by alcohol precipitation or vacuum drying.

7.3.2.9 FMRS Check 9: DNA denaturation check

Background: incomplete DNA denaturation will not provide reliable results. We assume the DNA denaturation was incomplete if: the ratio of the signal

intensity of the 96 fragment divided by the 92 fragment is lower than 03 and if: the signal intensity intensity of the 88 fragment divided by the 92 fragment is lower than 0.5. We assume that the DNA denaturation is partially incomplete if: the ratio of the signal intensity of or the 96 or the 88 divided by the 92 fragment is smaller than 0.5. If the ratio of the signal intensity of the 88 or 96 fragment is higher than 1.5 a warning is also given.
Problems are indicated when: in case the ratio of 88 control fragment and the 96 control fragment as opposed to the 92 fragment are both lower than 0.5, we assume that the denaturation completely failed and a warning is given. In case the denaturation failed 60 points are subtracted from the FMRS. In case the ratio of only the 96 or 88 as opposed to the 92 fragment is lower than 0.5 or higher than 2.5 a warning will be given and only 15 points are subtracted from the FMRS.
Recommendation with problems: in case there is a clear problem with the DNA denaturation, the reaction should be repeated and DNA should be denatured for at least 10 minutes at 96 degrees. In case the sample may contains high salt concentration it is advisable to desalt samples before repeating the reaction or diluting the sample by using lower sample volumes.

7.3.2 X and Y control fragments

Displayed as [X] & [Y]: Checks: If the X and Y control fragments were detected and if the signal intensity as opposed to the 92 control fragment was in the expected range. If the ratio of the signal of control fragment X as opposed to the 92-fragment signal is between ratios 0.2-3, the fragment will be marked green. In case the signal of X or Y control fragment was zero, the fragment will be marked red or as not present. In case the ratio is between 0-0.2 or, in case in of the X fragment higher than 3 and in case of the Y fragments a higher than 2, a warning will be given. The Y-control fragment is furthermore used to estimate the expected gender of each sample. Runs that have a Y-control fragment with a ratio higher than 0.15, as opposed to the 92 fragment, are expected to be males.

7.5 Using the fragment results explorer

By selecting “Open” from the right click menu on the fragments analysis settings window, while hovering above a sample row, you can open the fragment results explorer window. This can also be done by double clicking on the samples row on one of the QC icons. This fragments analysis explorer will allow you to examine each of the separate analysis steps of the

fragment analysis and also allows you to pinpoint more accurately where possible problems related to the fragment separation may have occurred. The fragment results explorer consists of 8 different tabs. The first tab contains the results of the separate factors that were used to calculate the FRSS and the FMRS (figure 29).

		P335-A2 lot 0809 BMP-ID 43-HYI-DNA Position:B2	
		Channel-1 6-FAM: probes (P335-ALL-IKZF1)	Channel-5 LIZ: size marker (GS500-250)
Size Marker	Correlation (R2)	n/a	0.9999
	Baseline size marker	n/a	-69
	Median signal height	n/a	598
	Maximum signal height	n/a	683
	Size marker signal sloping	n/a	-7.28
	Size marker signal widening	n/a	56.86
	Size marker complete	n/a	True
Probe Signal Check	Baseline probes	24.2	n/a
	Baseline probes percentage cut	26.91	n/a
	Primer %	14.28	n/a
	Median probe signal	2097	n/a
	Maximum probe signal	3415	n/a
	Probe signal sloping	48.42	n/a
	Probe signal widening	77.27	n/a
Found Probes	Found reference probes	10	n/a
	Found mutation probes	0	n/a
	Found Y probes	0	n/a
	Found total probes	51	n/a
	Percentage noise peaks	20.83	n/a
Found Control Fragments	Q-fragments median signal	0	n/a
	D-fragment 92 nt signal	981	n/a
	D-fragment 88 nt signal	1401	n/a
	D-fragment 96 nt signal	738	n/a
	X-fragment 100 nt signal	798	n/a
	Y-fragment 105 nt signal	1105	n/a
Control Fragment Results	Q-fragments median signal ratio	0	n/a
	D-fragment 88 nt ratio	1.43	n/a
	D-fragment 96 nt ratio	0.75	n/a
	X-fragment 100 nt ratio	0.81	n/a
	Y-fragment 105 nt ratio	1.13	n/a

Figure 29 *Fragment results explorer sample overview screen.*

Other available screens are:

1) Fragment analysis QC overview: This grid contains the results of all earlier discusses quality control factors. In case there is a problem with one of the separate factors the reason behind this can be found by hovering above the specific cell. These factors are also separately tested against thresholds and thereby also give the quality scores color indications in order to easily spot which factors were considered to be “bad”.
2) Raw data: displays the signals of the dye/data streams as your capillary electrophoresis device measures them. This screen can be used to

evaluate the quality of the fragment separation and to see if the dye filters are working correct.

3) Fragment profile: displays the baseline corrected signals of each dye/data streams that was set as a “probes” or “size marker” channel. Displayed signals of channels set as probe mix content will also show which signals were identified as peaks and what their relative length in nucleotides is. In this chart black triangle markers represent the position of the start of a peak, red circle markers represent the peak top and green asterisk markers represent the peak end. Above each peak top the estimated size called length is also displayed. By hovering over the peak top markers the tool tip information will appear showing the exact peak start, top and end data points, the peak height and the peak area. To make optimization of peak detection settings easier, the set minimal / maximum RFU and peak area% of the probes channels are displayed as line series.
4) Genomic profile: displays the baseline corrected signals of each dye/data streams that was set as a “probes” or “size marker” channel. Displayed patterns will also show which peak signals were identified as a probes, labels furthermore show the design length of each probe, gene name and exon number. The coordinates of the peak top of the peak that was recognized as the main peak related to a probe contains a green circular marker in case the probe is a reference probe and a purple circular maker in case it is a test probes.
5) Fragment analysis steps: on this page you view each of the fragment analysis steps by using the right mouse context menu. You can also navigate through the steps by using the key combination “ctrl” + “shift” + number key 1-10.
6) Binning: displays the peak heights on the y-axis and the estimated peak length of each detected peak on the x-axis. Next to this, using green and red stripes, the bin set that was used for data filtering is displayed. In case a peak signal was found that meets the requirements of the bin settings (as described in the CE devices chapter) and falls between the start and end length of a bin then that peak signal is assumed to originate from the probe product related to that bin and that signal will be called as that probe. On the X-axis right underneath each bin the probe name and rounded bin start-end length is displayed. In case a signal was related to a bin, then this bin will be colored green, in case no signals were related to that bin it will be colored red. By hovering over each bin the gene name, design length, bin start, center and end can be viewed in a tool tip. Next to this the median, average and standard deviation which are the result of the auto bin procedure are also displayed in the tool tip. Our algorithm is also able to link more than one peak to a probe within one sample. The amount of fluorescence of each

probe product may then be expresses the peak height, peak area of the main peak and the summarized peak area of all peaks in a bin.

7) File details: on this page you may view any file details that are added (mostly to ABIF) to files. Here you may view encoded data from each file allowing you to view details about the used capillary device and run settings. For example ABI gel type may be viewed by *GTYP-1*, Machine type by *HCFG-1 to 3*, injection time in seconds by *INSC-1,* injection voltage by *INVT-1,* capillary length by *LNTD-1,* run voltage by *LSRP-1*, capillary number by *LANE-1,* plate size by *96-Well,* run protocol by *RPRN-1*, used size standard by *STDF-1*, run temperature by *TMPR-1*, tube position by *TUBE-1* and user name by *USER-1*.

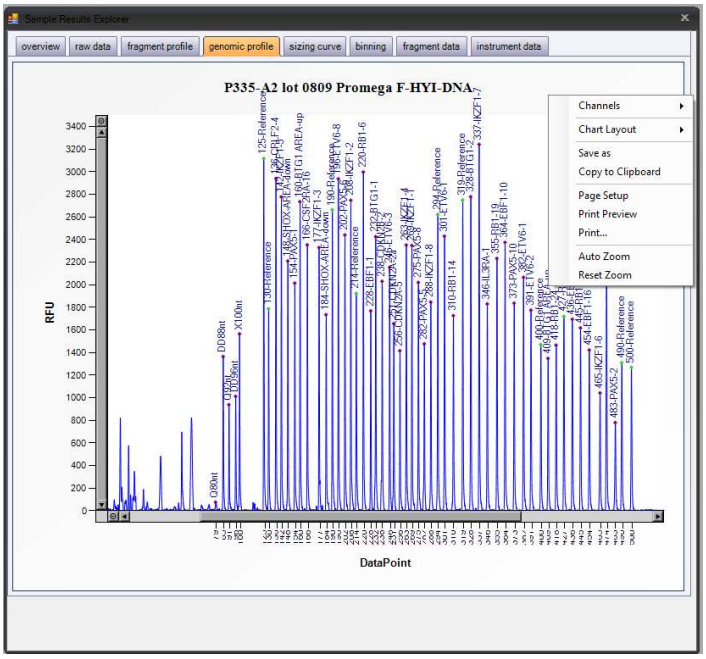


Figure 30 *Fragment results explorer genomic profile tab.*

Each tab of the fragment results explorer has several options that can be found in the right click menu (figure 30). You can for instance: view each channel independent, show or hide legends, save images in a wide variety of formats, copy to clipboard, print and make a print setup and use automatic zooming functions. Automatic zooming allows 3 levels of zoom, this being: show all detected peaks, show all recognized peaks and show all peaks

recognized as MLPA test probes. You may furthermore use manual zooming by clicking anywhere in the chart and dragging the mouse over the area you wish to zoom into. Exact details about control over charts and grid is described in the chapter *Context Menus*.

7.6 Creating a manual bin set for data filtering

Once all peaks have been size called, the profiles must be aligned to compare the fluorescence of the different targets across samples, an operation that is perhaps the single most difficult task in raw data analysis. Peaks corresponding to similar lengths of nucleotides may still be reported with slight differences or drifts due to secondary structures or bound dye compounds. These shifts in length make a direct numerical alignment based on the original probe lengths all but impossible. Our software uses an algorithm that automatically considers what the same peaks are between different samples, allowing easy peak to probe linkage, this procedure is called "*Auto binning*".
Our software algorithm follows four steps: reference profile analysis, applying and prediction of new probe lengths, reiteration of profile analysis and data filtering of all samples. The crucial task in data binning is to create a common probe length reference vector (or bin). In the first step our algorithm applies a bin set that searches for all peaks with a length closely resembling that of the design length of that probe. Next, the largest peak in each temporary bin is assumed to be the real peak descending from the related probe product. To create a stable bin, we calculate the average length over all real peaks of all used reference samples. If no reference samples exist, the median length over all collected real peak from all samples will be used. Since some probes may have a large difference between their original and detected length the previously created results may often not suffice. We therefore check if the length that we have related to each probe is applicable in our sample set. We do this by calculating how much variation exists over collected peaks length in each of the previous bins. If the variation was too large (standard deviation > 0.2) or no peak at all was found in any of the bins, the expected peak length for that probe will be estimated by prediction. The expected probe peak lengths may be predicted by using a second-order polynomial regression using the available data of the probes for which reproducible data was found. Even though a full collection of bins is now available, the lengths of the probe products that were predicted may not be very accurate. The set of bins for each probe in the selected MLPA mix will therefore be improved by iteration of the previous steps. The lengths provided for the bins are now based on the previously detected or predicted probe product lengths allowing a more accurate

detection of the real probe peaks. Probes that still were not found are predicted and a final length reference vector or bin is constructed for each probe. This final bin set can be used directly for data filtering but may also be edited manually in case the automatically created bin set may not suffice.
To edit the manual bin set right click on the fragments analysis experiment explorer and choose *"edit manual bin set*" from the context menu and then select the probe mix channel you wish to edit (see also figure 28).
Selecting this option will open the Coffalyser work sheet editor for manual bin sets allowing you to edit both the design an Coffalyser length that are used for the auto binning procedure but it also allows you to edit the manual bin set for that mix. The manual bin set allows you to set the start and end value in which a probe will be sought during data filtering. The manual filter values are loaded on default with the values of the Coffalyser length +/- 2 nt for the upper and lower bound. By selecting any sample in the left list box the sample will be loaded together with the detected peaks. In case a peaks fall within a bin and the signal of that peak met the criteria of the probe data filtering settings then the bin will be colored green, in case no peak was found on the peak did not match the criteria the bin will be red. This coloring method allows you to easily spot which bins should be changed. By selecting or changing any of the displayed bins, the displayed set in the chart will directly change into the set manual bin set, the color will also change into purple. By selecting a sample, you can view the data filtering results of that sample again, unless you enable the option box "*Always display the manual bin set while browsing through samples*". Enabling this option will hold on to the manual bin set as can be seen in the grid, allowing you to make changes and directly view if a peak actually falls within that newly created bin.
In the context menu under right click you may find a few options to make editing of a manual bin set more easy, from right mouse click menu select "set manual bin". This option will replace the manual bin set for either the selected row or for all rows with either: the design length, Coffalyser length, auto bin results for currently selected samples or auto bin results for the current experiment. The upper and lower bounds will be defined by taking the selected length and adding + / - the set search range of probes, as defined by the binning settings.

Figure 31 *Coffalyser work sheet editor - manual bin set*

Title	Coffalyser.NET analysis manual beta version
Status	Release candidate
Classification	Confidential
Versie	0.1

8. About the comparative analysis (copy number)

Signals of MLPA oligo-nucleotide probes are directly proportional to the amount of the target sequences present in a sample. Since these measurements have little meaning on itself, the signals of an unknown sample need to be compared to a reference in order to assess the copy number. For MLPA, in order to assess the copy number, signals of unknown samples can be compared to reference data by normalization. Normalization refers to the division of multiple sets of data by a common variable in order to negate that variable's effect on the data, thus allowing underlying characteristics of the data sets to be compared: this allows data on different scales to be compared, by bringing them to a common scale. The common variable or normalization constant thus needs to be derived from a factor that remains constant in each sample.
To make the normalization more robust, when normalizing the signal of a probe the procedure makes use of every MLPA probe signal, set as a reference probe to produce an independent ratio when comparing an unknown sample against a reference sample. The median of all produced ratios is then taken as the final ratio. This allows for the presence of aberrant reference probe signals without profoundly changing the outcome.
This process will then be repeated for each probe of each sample to each available reference sample, producing as many ratio results as there are reference samples. The final ratio will then estimated by calculating the average over these ratios. In case no reference samples are set, each sample will be used as reference and the median over the ratios be calculated.
During the normalization the software also calculates the average, median and the standard deviation (reproducibility) of each probe over sample results that have the same sample type in the performed experiment. Reference samples are assumed to be genetically equal, so the effects of sample-to-sample variation on the inidividual probe ratios can then be estimated by the reproducibility of these results. This data can later also be used for sample profile comparison of unknown samples to the results over a group of samples (e.g. a set of positive or negative reference samples).
In order to get an estimation of the reproducibility of each independent unknown sample probe ratio result, the algorithm combines the variation found over the set reference samples, with the discrepancies computed between the probe ratios per reference probes within the sample.
This works as following: during normalization our algorithm makes use of each reference probe for normalization of each test probe; thereby

producing as many dosage quotients (DQ) as there are references probes. The median of these DQ's will then be used as the definite ratio. The median of absolute deviations (MAD) between the computed dosage quotients may therefore reflects the introduced mathematical imprecision of the used normalization factor.
By combining the standard deviation found over the reference sample data with this MAD factor and mulitplying it by two we can estimate a 95% confidence range for a probe result.
By comparing each sample's test probe ratio and its 95% confidence range to the available data of sample groups in the experiment, we can conclude if found results are significantly different from e.g. the reference sample population or equal to a positive sample population.
The algorithm then completes the analysis by evaluating these results in combination with the familiar set of arbitrary borders used to recognize gains and losses. A probe signal in concluded to be aberrant to the reference samples; if a probe signal is significantly different as from that reference sample populations and if the extent of this change meets certain criteria. The results are finally translated into easy to understand bar charts (figure 2) and sample reports allowing users to make a reliable and astute interpretation of the results.

Data signal to size sloping

An effect that is commonly seen with MLPA data is a drop of signal intensity that is proportional with the length of the MLPA product fragments. This signal to size drop is caused by a decreasing efficiency of amplification of the larger MLPA probes and may be intensified by sample contaminants or evaporation during the hybridization reaction. Signal to size drop may further be influenced by injection bias of the capillary system and diffusion of the MLPA products within the capillaries.
In case the drop in signal is equal between each unknown sample and reference sample no problem exist, because this effect will be normalized out of the equation. However when a difference in this extent exists, results may be biased. In order to measure and if needed correct for this, Coffalyser.NET follows several steps.

1) Normalization of all data in population mode. Each sample will be applied as a reference sample and each probe will be applied as a reference probe.
1) Determination of significance of the found results by automatic evaluation using effect-size statistics and comparison of samples to the available sample type populations.

2) Measure of the relative amount of signal to size drop. If the relative drop is less than 10% a direct normalization will suffice, any larger drop will automatically be corrected by means of regression analysis (step 4-5).
3) Before correction of the actual amount of signal to size drop, samples are corrected for the MLPA mix specific probe signal bias. This can be done by calculating the extent of this bias in each reference run by regressing the probe signals and probe lengths using a least squares method. Correction factors for these probe specific biases are then computed by dividing the actual probe signal through its predicted signal. The final probe-wise correction factor is then determined by taking a median of the calculated values over all reference runs. This correction factor is then applied to all runs to reduce the effect of probe bias due to particular probe properties on the forthcoming regression normalization.
4) Next we calculate the amount of signal to size drop for every sample by using a function where the log-transformed probe bias corrected signals are regressed with the probe lengths using a specialized local median least squares method. Signals from aberrant targets are left out of this function, by applying an outlier detection method that makes use of the results found at step 2. The signal to size corrected values can then be obtained by calculating the distance of each log transformed pre-normalized signal to its predicted signal.
5) Normalization of signal to size corrected data in the user selected mode (usually block method using reference probes) and determination of significance of the found results.

Even though each of these different steps has default settings for analysis, most steps may be adapted in order to provide the possibility to optimize for specialized data types.

8.1 Setting up the comparative analysis

After you have analyzed and explored your fragment data, you can navigate to the next step, which is the sample dependent comparative analysis. Since the fragment analysis is sample independent you may select any combinations of samples in the fragment analysis. Please note that leaving out samples, may influence the normalization of all samples and thus the probe ratios of all samples.

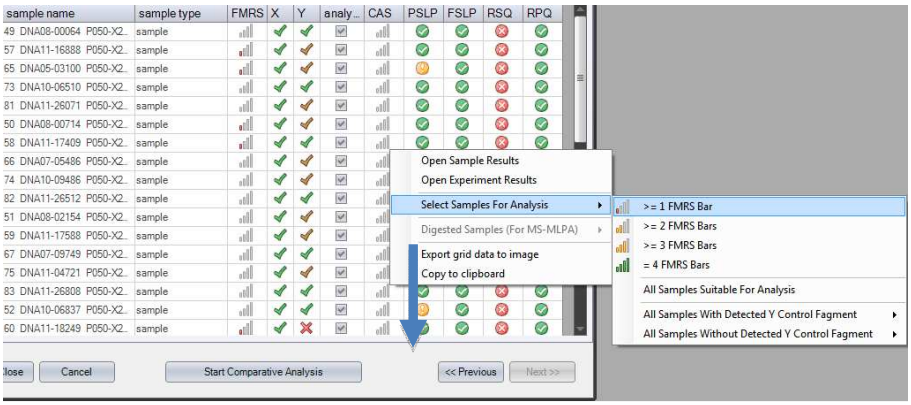


Figure 32 *Comparative analysis sample selection menu*

The first thing that needs to be done at the comparative analysis tab is the selection of samples that will be included in the normalization. To make the selection easier you may use the right click menu to make a pre-selection of samples based on their FRMS score. Right click anywhere in the grid, select the option; “Select samples for comparative analysis”. Next select a level of quality you which to apply for the comparative analysis. Dependent on the setting of the study, e.g. research or diagnostic a higher quality level may be desired. You can further adjust the selection of samples by using the option box in column “analyze”. After finishing your selection of samples click on the button “Start comparative analysis” (blue arrow, figure 32), which will open the comparative basic analysis settings form (figure 33).

Basic normalization settings:

On this form you can adjust some of the settings that will influence some of the most basic analysis settings. On default all settings are set to auto, resulting in a multistep analysis where the best settings are chosen dependent on: the number of samples and their sample types, the MLPA mix, presence of reference probes and results obtained from earlier steps during the analysis. By using the different tabs users may influence the parameters for the different steps of the comparative analysis. On the first tab we find the basic normalization settings.

1) Normalization metric: the normalization metric is the system of measurement of each detected probe that will be used during normalization. If this option is set to “Calculate best (signal to noise)

[default]" each possible probe metric will be compared to each other, and the metric showing the highest signal as opposed the amount of noise will be used for normalization. Users may furthermore choose if they want to use: peak heights, peak areas, or peaks areas including their siblings are used for normalization. Peak areas plus siblings means that all peaks that passes the minimal peak detection thresholds and fall within the bin set of a probe are summarized and used for normalization.

2) Normalization factor (intra): during normalization of a test sample against a reference sample each test probe will be normalized using each set reference probe thereby producing as much ratios as there are reference probes. To create a final estimator (dosage quotient) for each test probe the "normalization factor (intra)" will be taken over these ratios. On "auto [default]" this factor is set to "median", thereby allowing some of the reference probes to be altered (<40%) without having an effect on the final results. User may however also choose to use the average, minimum of maximum of the collected ratios. Minimum and maximum should be avoided unless you are choosing this for a special kind of analysis.
3) Normalization (inter): in the presence of multiple reference samples each test sample will be compared to each reference sample, thus generating as many dosage quotient or ratios as there are references samples for each probe. In order to obtain a single result for each sample probe a for these dosage quotients the "normalization factor (inter)" will be taken over these ratios. When this option is set to auto the average will be taken when there are more than 2 reference samples present, if no reference samples are available all samples will be used as a reference and the median will be taken. User may however also choose to use the minimum or maximum, which should only be chosen if you are choosing this for a special kind of analysis.
4) Arbitrary ratio border (low/high): the arbitrary borders are the set borders where we expect normal results to fall in between. In figure 32, a delta of ratio 0.3 is set as opposed to the reference (which is always 1), resulting in a normal range of ratio 0.7-1.3 for results that appear to be normal or equal to the signals found in the reference samples.

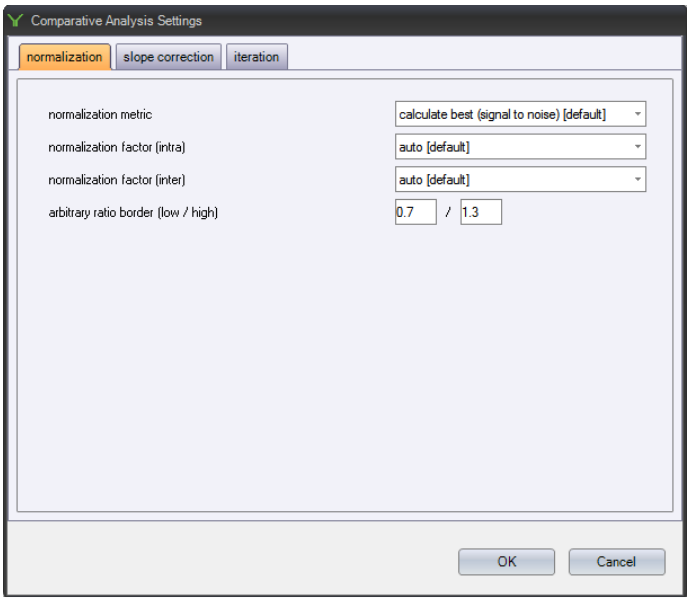


Figure 33 Basic comparative analysis normalization settings.

Slope correction settings:

The second tab contains the analysis details concerning the slope correction of the data. Slope correction of data may be necessary in case there is too much difference in the signal to size drop between reference samples and test samples. A difference in signal to size sloping may cause the ratios of the shorter probes seem to be gained and the ratios of the longer probes may seem to be losses while this is actually caused by a difference in fidelity of the polymerase between the reference sample PCR reaction and the unknown sample PCR reaction. In case the difference in signal to size drop is minimal, no slope correction is necessary and we also recommend it in such cases since regression analysis is much more sensitive as compared to regular normalizations. This signal to size drop is caused by a decreasing efficiency of amplification of the larger MLPA probes and may be intensified by sample contaminants or evaporation during the hybridization reaction. Signal to size drop may further be influenced by injection bias of the capillary system and diffusion of the MLPA products within the capillaries. You can change several settings in order to optimize the slope correction procedure.

1) Slope correction: slope correction aims to correct the drop in fragment signal intensity to the length of each fragment that is unrelated to the number of target sequences available in each sample. When this option is set to: “auto (if>10%) [default]”, slope correction will only occur if the difference in sloping between reference and samples is more than 10%. If this difference is less than 12% sloping correction may not be required because the normalization itself will then resolve this issue. You can furthermore choose to always do the slope correction or never.
2) X metric: main metric that will be used for the regression analysis's on the X-axis. For each probe signal we can apply either the lengths or data points related to the probes.
3) Y metric: by changing the Y-metric you can influence whether the raw signals will be corrected or the pre-normalized ratios. Instead of the correcting the signals, the pre-normalized ratios calculated in the first normalization of all data in population mode may also be corrected and normalized afterwards.
4) Log correction of signals: determines whether or not the signals used for regression analysis are first converted to a log scale before creating the regression line.
5) Major outlier filter (high / low): this first outlier filter is used to ignore signals based on the pre-normalized ratios. By setting a very rude filter you can ignore signals that are very aberrant, this will help with better fitting of regression lines.
6) Ignore major outlier filter (for dynamic detection): determines whether or not the probes that were detected, as major outliers should be left out the regression line dynamic detection method (see outlier detection method).
7) Outlier detection method: determines the way how outlier signals (probes) should be determined before plotting a regression line through the signals. Note! Outlier detection methods should only be applied on regression lines of the type "Least squares" or polynomial. The local linear method and least squares local median are methods that already ignore outlier by their methodology and extra outlier detection may make these methods less robust.

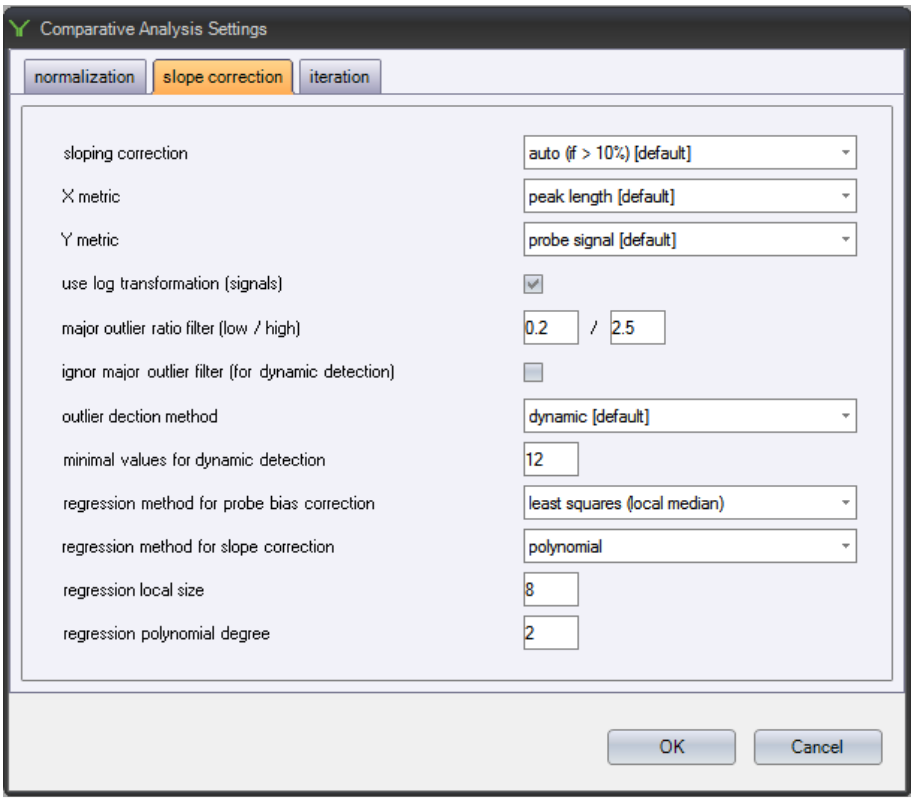


Figure 34 Sloping correction analysis settings.

Iteration settings:

The final tab contains settings concerning the iterations steps of the comparative analysis (figure 35). Iteration means the act of repeating a process usually with the aim of approaching a desired goal or target or result. Each repetition of the process is also called”iteration," and the results of one iteration are used as the starting point for the next iteration. Coffalyser.NET allows repeated normalization rounds where the used reference probes, reference samples and slope correction methods may be optimized based on the results found in the previous rounds. During these rounds of normalization we aim to make the results "more normal" or perform normalization more towards the original set normal image. Results will thus always be normalized to the set median or average reference sample / reference probe status.

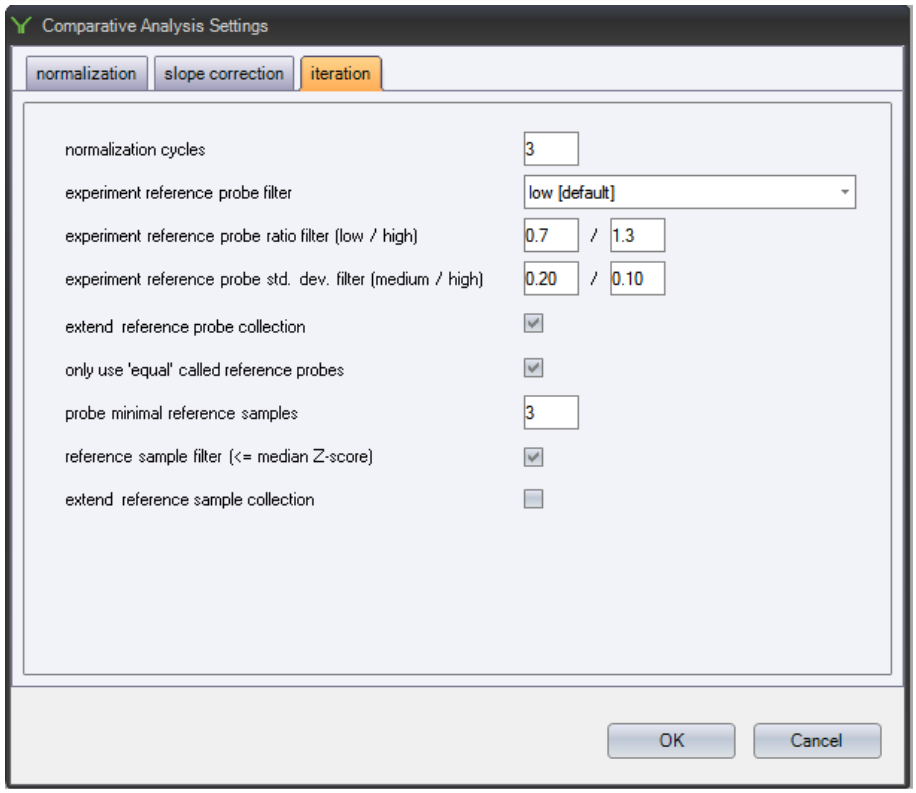


Figure 35 Iteration normalization analysis settings.

On default settings the number of iteration round is set to 1, which means a single round of analysis without further adjustment of settings. To use the iteration, the number of rounds need to be at least 2, and in most cases when using just 3 rounds the iterations is optimal. You can furthermore change several settings in order to optimize the iteration procedure.

1) Normalization cycles: the normalization cycles refer to optional experimental iteration of results. Iterative normalization means that all samples will be completely analyzed where after the results will be automatically interpreted and a new normalization starts with new parameters based on the results of the previous normalization. This method allows a number of methods, which are discussed more extensively in the advanced analysis section. In short, each sample may obtain sample related reference probes and reference samples, which were found to be normal or equal in the previous analysis. This method works best in case you have a large sample collection and no reference probes and you do not have any background information about the samples.
2) Experiment reference probe filter: this filter adapt the reference probes influences the way the reference probes are selected in the next round of normalization. The filter uses the statistical results as found over the combined samples of the types reference sample and (test) sample. Depending on the type of filter the effect of each settings should be combined with the "Experiment probe reference filter (low/high)" and the "Experiment reference probe std. dev. filter (medium / high)". In case the filter is set to low, medium, high or incremental; the probes that have an average ratio, as calculated over the reference samples or over all test samples, that is outside the "Experiment probe reference filter" will NOT be used as reference probes. In case the filter is set to medium; the probes that have a standard deviation, as calculated over the reference samples or over all test samples, that is higher than the "Experiment reference probe std. dev. filter" medium value (0.2 at default) will NOT be used as reference probes. In case the filter is set to high the same rule applies but now the standard deviation is compared the maximal values set under high (0.1 at default).
3) Extend reference probe collection: in case this option is enabled the selection of reference probes is extended to all probes that pass the criteratia at point 2. If this option is off, the criteria will only be applied at the reference probes that are set in the first round of normalization, this selection is dependent on the used analysis method. If the analysis method was set to block then all selected reference probes in the active sheet were used; if the analysis method was set to population then all probes were used as reference probes.
4) Only use "equal" called reference probes: this option limits the use of the earlier selected reference probes to those that were earlier found to be equal to the reference samples collection. What the criteria are for a probe to be equal to the reference sample collection is explained further down in this chapter. In short probes that are equal to the reference

sample collection are fall within the 95% confidence range of this population and do not cross the arbitrary set borders (default 0.7-1.3).

5) Probe minimal reference samples: this settings defines the minimal number of reference samples that should be left in the end of the analysis per probe. In most cases the set reference samples will be equal for each sample, however when using the options "only use "equal" called reference probes and "reference sample filter [<= median Z-score], the used reference sample and probes may be different for each sample and a minimal number of reference samples that should remain is recommended.
6) Reference sample filter [<= median Z-score]: this option enables users to minimize the reference sample collection by decreasing the used reference sample signals each round by half. If we for instance start with 10 samples and no reference samples the first analysis will use all samples as reference samples and the final estimator for each ratio will be estimated by taking the median. By applying the Z-scores the reference samples will be limited to the signals that had a Z-scores that is lower than the median Z-scores overall samples divided by two. This basically minimizes the used signals to the 50% that are closest around the original reference set. By increasing the number of cycles the number of used reference samples will be divided in two, each round, until the minimum number of reference samples is reached.
7) Extend reference sample collection: extending your reference sample collection means that we can use the data of all samples in order to create a new reference sample collection. This option only has use in case you are already using a collection of reference samples but you want to increase this set automatically. It should be noted that this option is OFF on default, because it may skew the results.

After changing the desired settings click on “OK” to start the analyzed. Dependent on the number of samples, the number of reference samples, the analysis settings and the composition of your computer the analysis may take as little as 10 seconds while large experiment may take several minutes.

8.2 About the comparative analysis quality scores

After the analysis is finished you will be confronted with a number of quality scores that may indicate the quality of the normalization, slope correction and overall analysis quality of each sample (figure 36). Evaluation of the comparative analysis quality scores should be done as described in 7.3. Here we describe the meaning of the different displayed scores.

8.2.1 PSLP

Check 1: Pre-normalization signal sloping probes displays the relative amount of signal to size drop of the probe fragments of a sample as opposed to the reference. See also 7.3.4 FMRS Check 4: Signal drop of the internal run of the probe fragments.

8.2.2 FSLP

Check 2: Final-normalization signal sloping probes displays the relative amount of signal to size drop of the probe fragments of a sample as opposed to the reference after the signal have been corrected for signal to size sloping effects. See also 7.3.4 FMRS Check 4: Signal drop of the internal run of the probe fragments. This measurement checks if performed slope correction method was successful.

8.2.3 RSQ

Check 3: Reference sample quality displays if relative probe signal inconsistencies existed in the selected reference sample population. The amount of variation is estimated by measuring the standard deviation over the calculated final normalized dosage quotients of each probe over all the reference samples.

8.2.4 RPQ

Check 4: Reference probe quality displays if relative reference probe inconsistencies existed in the complete sample population. The amount of variation is estimated by measuring the standard deviation over the calculated ratios which are generated when a probe is normalized against each separate reference probe during each sample to reference normalization.

8.2.5 CAS

Check 5: Coffalyser analysis score displays the quality of the complete analysis of a sample that comprises all quality points calculated during the fragment and comparative analysis into a single score.

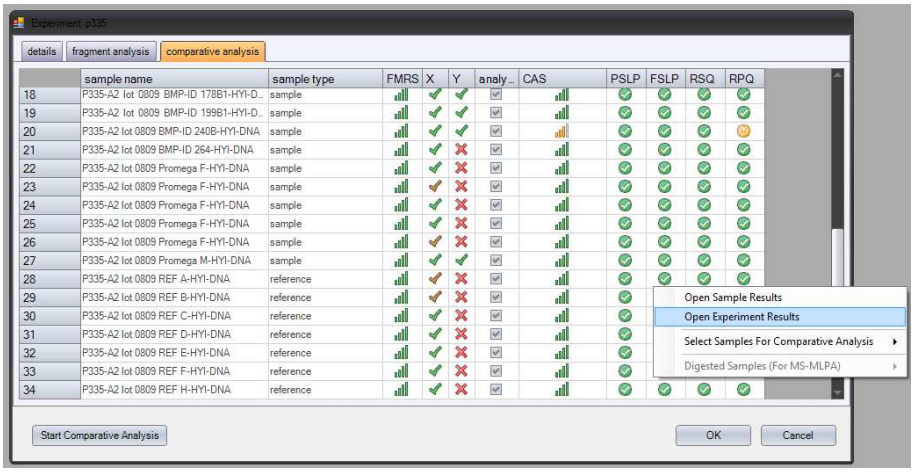

Experiment: p335
details
fragment analysis
comparative analysis
sample name
sample type
FMRS
X
Y
analy...
CAS
PSLP
FSLP
RSQ
RPQ
18 P335-A2 lot 0809 BMP-ID 178B1-HYI-D... sample
19 P335-A2 lot 0809 BMP-ID 199B1-HYI-D... sample
20 P335-A2 lot 0809 BMP-ID 240B-HYI-DNA sample
21 P335-A2 lot 0809 BMP-ID 264-HYI-DNA sample
22 P335-A2 lot 0809 Promega F-HYI-DNA sample
23 P335-A2 lot 0809 Promega F-HYI-DNA sample
24 P335-A2 lot 0809 Promega F-HYI-DNA sample
25 P335-A2 lot 0809 Promega F-HYI-DNA sample
26 P335-A2 lot 0809 Promega F-HYI-DNA sample
27 P335-A2 lot 0809 Promega M-HYI-DNA sample
28 P335-A2 lot 0809 REF A-HYI-DNA reference
29 P335-A2 lot 0809 REF B-HYI-DNA reference
30 P335-A2 lot 0809 REF C-HYI-DNA reference
31 P335-A2 lot 0809 REF D-HYI-DNA reference
32 P335-A2 lot 0809 REF E-HYI-DNA reference
33 P335-A2 lot 0809 REF F-HYI-DNA reference
34 P335-A2 lot 0809 REF H-HYI-DNA reference
Open Sample Results
Open Experiment Results
Select Samples For Comparative Analysis
Digested Samples (For MS-MLPA)
Start Comparative Analysis
OK
Cancel

Figure 36 *Comparative analysis quality score screen*

8.3 About the comparative experiment results explorer

Coffalyser.NET provides two ways to evaluate the results: exploration of the results of the complete experiment or exploration of results of a single sample. To open the experiment explorer: right mouse click on the grid showing the quality scores and select from the right click menu “Open experiment results”. The comparative analysis experiment explorer has three tabs allowing getting a quick overview of the results of the complete experiment.

8.3.1 Comparative analysis experiment explorer statistical overview chart

The last tab shows a statistical overview chart which loads with the statistical results found over all samples that were set with the sample type “sample” (figure 37). All probe results are displayed as ratios on the Y-axis, the X-axis will on default load on displaying the map view locations of the target sequences of the probes obtained by the hg18 tracks generated by UCSC and collaborators worldwide. The labels above the probes on default load with a text field containing “probe length - gene name of target sequence – exon number within gene of target sequence”, e.g. “126 – DMD – 01”, which thus suggests that this probe had a design length of 126 nucleotides and was targeted to exon 1 of the DMD (dystrophy) gene. The different vertical stripes or color bands indicate which probes fall within a certain region. On default the chart will load placing all probes within one region that are located on the same chromosome arm. Other regions include: chromosome,

chromosome band or user defined regions. On default information on user defined regions are filled in by MRC-Holland. All results are furthermore organized according to the MRC-Holland recommended order. In practice this often means that test probes and reference probes are separated and sorted by the hg18 tracks, if no recommended order exists results will automatically organized by the hg18 tracks.

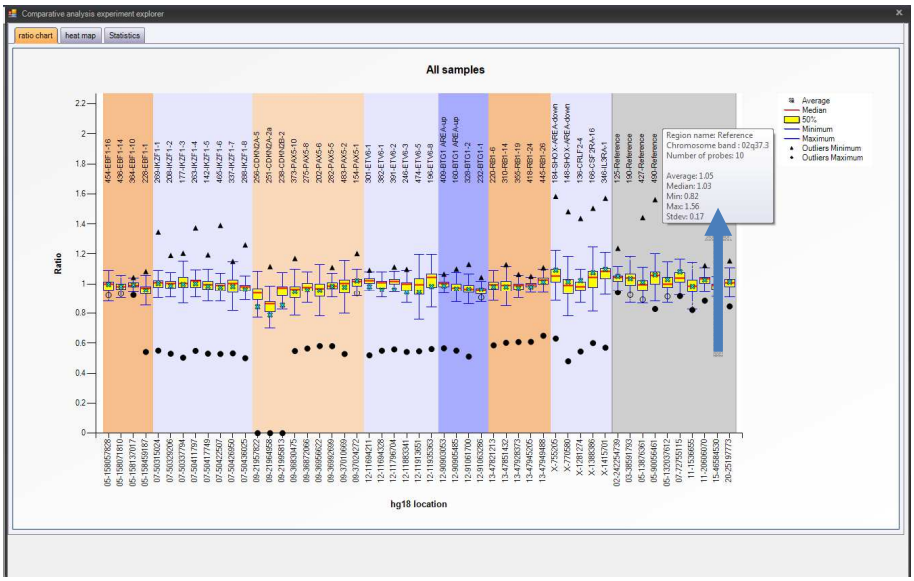


Figure 37 *Comparative analysis experiment explorer statistical overview chart*

Results of all samples of the same sample type for each probe are displayed by a box plot (also known as a box-and-whisker diagram or plot) graphically depicting the results in groups of numerical data through their five-number summaries: the smallest observation (theoretical minimum), lower quartile (Q1), median (Q2), upper quartile (Q3), and largest observation (theoretical maximum). A box plot may also indicate which observations, if any, might be considered outliers. The quartiles of a set of values are the three points that divide the data set into four equal groups, each representing a fourth of the population being sampled. IQR is the distance of Q1 to Q3 thus containing 50% of all values, which is depicted in the chart by the yellow box. The theoretical minimum is then estimated by Q1 minus 1.5xIQR and the theoretical maximum by Q3 plus 1.5xIQR, which are displayed respectively as the lower and upper whiskers. If results exist that fall outside the range of the theoretical minimum and maximum they will be displayed as by black

round markers for the minimum and black triangle markers for the maximum values. The average is depicted in the chart by a blue cross and the median by a red stripe.
Whenever the mouse hovers above any of the displayed symbols extra information will be displayed by a tooltip box. This allows you to quick find the exact result numbers and additional information about the probe and its target sequence. The right click context menu enables you to customize the chart completely (figure 38).

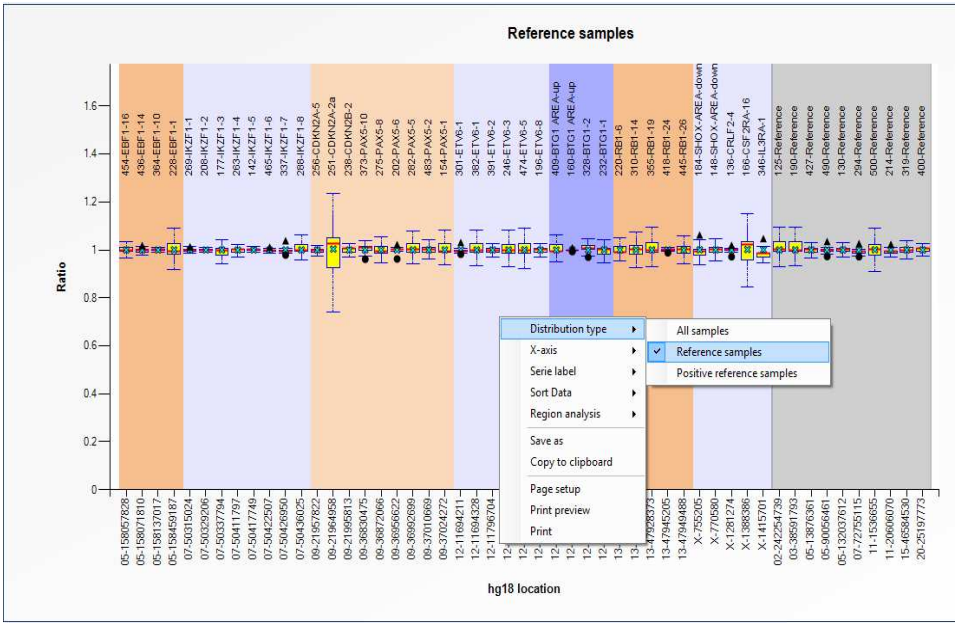


Figure 38 *Comparative analysis experiment explorer statistical overview chart showing descriptive statistics of the reference samples*

The following features may be selected:

1) Distribution type: changing the distribution type results in the display of the descriptive statistics of all samples of a sample type. By displaying results of the “reference samples” you can easily evaluate the reproducibility of each probe in that experiment, assuming the reference samples were genetically equal and the reference samples were properly dispersed through the experiment (figure 32).

2) X-axis: by selecting a different field for the X-axis you may display: “probe length - gene name - gene exon number”, “gene name - gene exon number”, hg18 track location, chromosomal band or probe length.
3) Series label: allows display of the same info described at point 2.
4) Sort data: change the order of the probes sort either by: recommended order, hg18 tracks, chromosomal band or probe lengths. Sorting on probe length allows you to see if there was a general trend of signal sloping within this data set.
5) Region analysis: changing the region analysis allows you to change what probes are harboring together in a stripe. This requires some calculation since you may also find a statistical description of all probe results that fall in that region by hovering above any of the stripes or regions (blue arrow, figure 31).
6) Other options include the option to: save images in a wide variety of formats, copy to clipboard, print, print setup and print preview
7) Channels: enables or disables the results of the data found for each channel that was set as a probe mix channel. This option works only in multichannel modus, also see the advanced analysis section.

8.3.2 Comparative analysis experiment explorer heat map grid

The second tab shows the ratio results of all samples for each probe in a sorted grid. The probes are displayed on the rows while the columns may contain a number of hierarchies. Hierarchical levels allow you to hide or show information by clicking on the plus sign in the left top of the column headers. You may for instance click on the plus sign of the column header or double click anywhere in the header that states “probe target info”, which will then open the columns: probe name, chromosomal position, hg18 track position, probe length and the recommended order (blue arrow, figure 39). Each of these columns can be used to sort the whole grid by clicking on the column header cells.

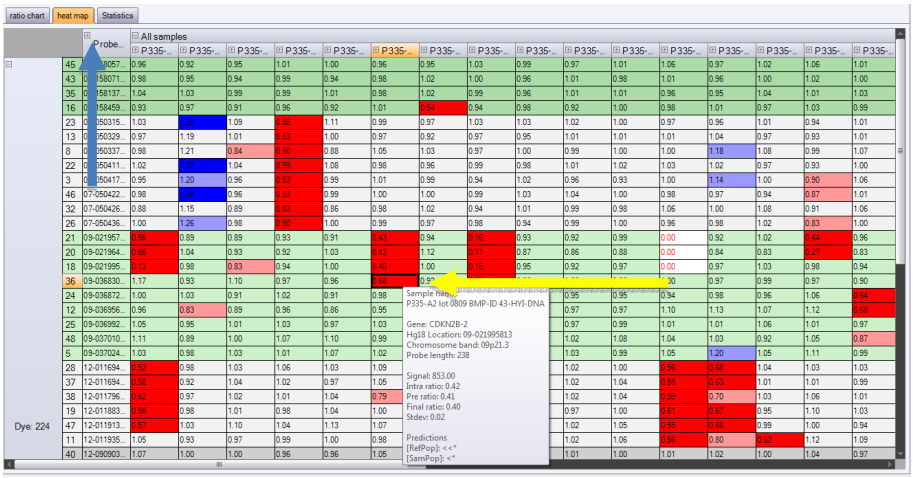


Figure 39 *Comparative analysis experiment explorer heat map grid overview*

The top levels of the columns contain a maximum of 4 entrees: probe target info, all samples, reference samples and positive samples. Each sample type group then contains all samples underneath it which levels are already opened on default. Each sample can be furthermore opened to display separate information about each detected probe peak (figure 40). The levels underneath each sample are closed on default and can contain the following levels: peak signal, intra normalized ratio, pre-normalized ratio, ratio without iteration, final ratio, standard deviation and distribution comparison results against the collection of test samples, reference samples and positive samples. This information may also be summoned by hovering above a cell in the grid; a tool tip control will then provide all available data for that result (yellow arrow, figure 40). For more information about these different normalized ratios and distribution comparisons values also see the FAQ in the end of this document or published articles about the methodology behind Cofaflyser.NET (J. Coffa, 2011; J. Coffa 2008).

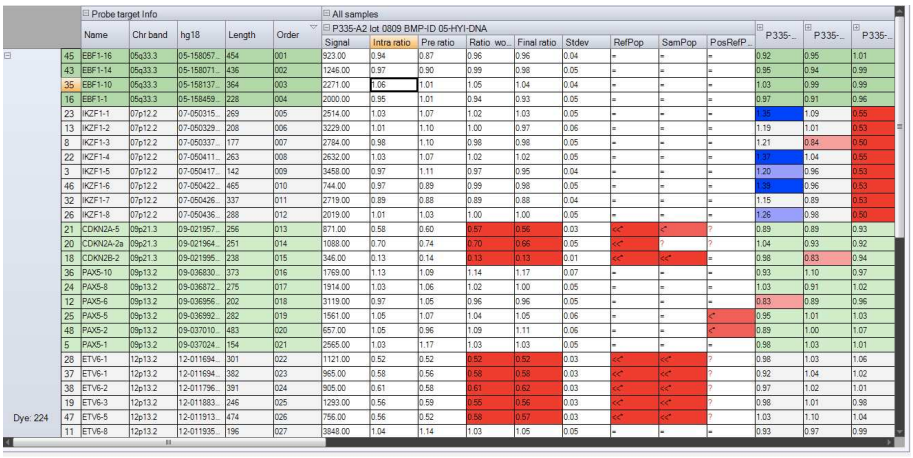


Figure 40 *Comparative analysis experiment explorer heat map grid overview with probe target info columned opened and sample result lower levels also opened.*

The probes in the grid are on default sorted by the recommended order; if this does not exist the hg18 tracks will be used for data sorting. Each row or probe is related to a certain region which, depending on the settings, will group probes together by giving them a certain color. On default probes that have their target sequence to the same chromosomal arm are grouped together. Cells that contain probe ratio results can be colored in different ways depending on the set conditional format. On default cells will be colored if they were found to be different from the used reference sample collection.

- Blue or “>>* “(figure 45c): Cells will be color blue if found results meets 2 criteria. First the magnitude of the probe ratio exceeded the upper set arbitrary border value (on default 1.3). Secondly the 95% (2 standard deviations) confidence range of the probe did not overlap with the 95% confidence range of that probe in the reference sample population. In the lower levels of the distribution comparison to the reference sample population this result will be noted by the symbol “>>*”.
- Blue one tint lighter or “>*”: If found results, did not meet the criteria of 1 then 2 new criteria are tested which if realized will color the cells one tint lighter blue. First the magnitude of the probe ratio exceeded the upper set arbitrary border value (on default 1.3). Secondly the 68.1% (1 standard deviation) confidence range of the probe did not overlap with the 68.1% confidence range of that probe in the reference sample

population. In the lower levels of the distribution comparison to the reference sample population this result will be noted by the symbol “>*”.

- Blue two tint lighter or “>>”: If found results did not meet the criteria of 1 or 2 then we will check if found probe results has a significant different to the reference sample population without employing the magnitude of the probe ratio. Cells will be color blue 2 tint lighter blue if the 95% confidence range of the probe did not overlap with the 95% confidence range of that probe in the reference sample population. Probe results from samples with mosaic cell population may often be contaminated with normal cells which may cause the magnitude of the probe ratio to be within the set of arbitrary border, while the result may still be significantly different from the reference population. In the lower levels of the distribution comparison to the reference sample population this result will be noted by the symbol “>>”.
- Region color or white with black text or “=” (figure 45a): The color of the cells will not change if the result was found to be equal to the reference sample population. Results are assumed to be equal if 2 criteria are met. First the magnitude of the probe ratio falls within the lower and upper set arbitrary border values. Secondly the probe result falls within the 95% confidence range of that probe in the reference sample population. In the lower levels of the distribution comparison to the reference sample population this result will be noted by the symbol “=”.
- White with red text or “?” (figure 45e & 39f): : The cells will become white with red text if the result was found to be ambiguous. Results are assumed to be ambiguous if the magnitude of the probe ratio falls does not fall within the lower and upper set arbitrary border values. The result was however also found to fall within the 68.1% confidence range of that probe in the reference sample population. This indicates that this probe was found to be very variable in the reference sample collection and no unequivocal conclusion can be taken from this result. In the lower levels of the distribution comparison to the reference sample population this result will be noted by the symbol “?”.
- Red or “<<*” (figure 45d): Cells will be colored red if the found result meets 2 criteria. First the magnitude of the probe ratio is lower than the lower set arbitrary border value (on default 0.7). Secondly the 95% (2 standard deviations) confidence range of the probe did not overlap with the 95% confidence range of that probe in the reference sample population. In the lower levels of the distribution comparison to the reference sample population this result will be noted by the symbol “<<*”.
- Red one tint lighter or “<*”: If found results did not meet the criteria of 6 then 2 new criteria are tested which if realized will color the cells one tint

lighter red. First the magnitude of the probe ratio exceeded the upper set arbitrary border value (on default 0.7). Secondly the 68.1% (1 standard deviation) confidence range of the probe did not overlap with the 68.1% confidence range of that probe in the reference sample population. In the lower levels of the distribution comparison to the reference sample population this result will be noted by the symbol “<*”.

- Red two tint lighter or “<< (figure 45b)”: If found results did not meet the criteria of 6 or 7 then we will check if found probe results has a significant different to the reference sample population without employing the magnitude of the probe ratio. Cells will be colored 2 tint lighter red if the 95% confidence range of the probe did not overlap with the 95% confidence range of that probe in the reference sample population. . In the lower levels of the distribution comparison to the reference sample population this result will be noted by the symbol “<<”.
- Yellow with red text or "<<**": if a probe could not be related to any peak then the cell will be colored bright yellow. It's always recommended to confirm these kind of results by evaluation of the raw electropherogram, to ensure that the signal is actually gone.

By using the right mouse click menu you may find the following options:

1) Open sample results: this will open the sample explorer (explained in the next chapter) allows a more detailed exploration of all data on this sample.
2) Region analysis: changing the region analysis allows you to change what probes are sorted together and obtain the same default cell color. Applicable region settings are chromosome, chromosome arm, chromosome band or custom region numbering.
3) Show data type: this option allows you to change the data of the main grid. You can select RNA, to display the intra-normalized ratios. It should be noted that this option is only useful when doing an RNA analysis. It is then possible to view the Intra-normalized ratio get see how each signal relates proportionally to the set reference probes. Alternatively, DNA can be selected, which will display the ratio as they are normalized against the reference samples. When a copy number / methylation status analysis is performed, either DNA results, MS results or both may be viewed at the same time. On default both copy number and methylation status are loaded and placed directly next to each other.
4) Conditional format: changed the conditional method of the grid result cells. Other than the earlier described default color-coding cells may also be colored by: 1. Arbitrary borders, only comparing the magnitude against the set arbitrary borders. Cells that have a probe ratio result higher than the upper border will be red, cells that have a result under

than the lower border will be blue. 2. Gradient, comparing probe ratios against an array of arbitrary borders with steps of 0.1 where each value that is higher than 1 becomes more blue while each value lower than 1 becomes red. 3. Hierarchical heat map, gives a cell a color based on the rank of that result in either the results in that same sample, the results of all sample of the same type or the results of all samples. 4. Probability scores, cells will be colored based on their population comparison value as described earlier. Results may however also be compared to the test sample population or positive sample population.

5) Resize column width: by resizing the column widths you may fit more or less sample data on your screen. Columns widths may be changed directly to 25, 50, 100 or 150 points or may be auto resized to fit the sample name of each column. Alternatively the width can be changed gradually by increasing or decreasing with steps through the menu or by using the short cut CTRL + SHIFT + Plus or Minus key.
6) Resize row height: by resizing the row height you may fit more or less probe data on your screen. Row heights may be changed directly to 8, 10, 15 or 20 points. Alternatively the height can be changed gradually by increasing or decreasing with steps through the menu or by using the short cut CTRL + ALT + Plus or Minus key.
7) Export grid: exports the displayed grid data to a file in *.*csv, *.*HTML, *.*XML document or *.*XML spreadsheet format.
8) Export pdf overview: creates a quality control list of each sample with their quality control scores (see chapter 7.3 and 8.2) in pdf format (figure 41) and creates a pdf overview of all samples final ratio results (figure 42) . In this document result cells become bold if they do not fall within the set arbitrary border and obtain one or two asterisk symbols behind the value is the result is different than 68.1% or 95% respectively different from the reference sample population.
9) Hide column: by selecting this option you can hide any column in the grid. When hiding columns the effect will be passed through on the export grid function, but not on the PDF reports.
10) Show all columns: after hiding columns all invisible columns can be recovered by selecting this option.
11) Channels: enables or disables the results of the data found for each channel that was set as a probe mix channel. This option works only in multichannel modus, also see the advanced analysis section.

PDF experiment ratio overview

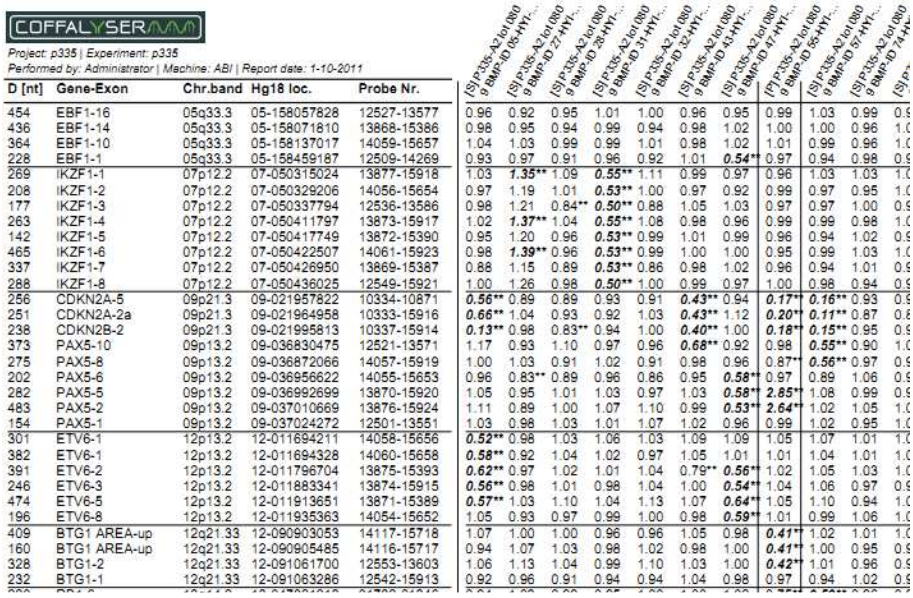
Coffalyser.NET allows a number of pdf report functions which create easy storable files that show a complete overview of either a single sample or the complete experiment. By selecting the export PDF function from the

experiment results explorer, a document will be created that has 3 subsections. The first section contains an overview of all samples, their analysis status and all other important quality control aspect that were measured during the analysis. The second section contains a print of the ratio overview grid in PDF form; here you can find each sample with their probe results. In case the sample's quality was below standard, the name will be bold. The data in this grid will be sorted as set in the experiment results explorer. Ratios of probes that fall outside of the arbitrary borders are depicted with a bold font; ratios of probes that were found to be statistically different from the reference sample population are depicted with a single apteryx in case they fall outside of 68% of the population and two apteryxes in case they fall outside of 95% of the population. Finally the last page of the report contains the statistical measurements over each sub population of samples. We measure the average, median, minimum, maximum, standard deviation and the median of absolute deviations (MAD) over the reference samples, test samples and positive reference sample for both copy number and methylation status.

Sample QC list Project: p335 | Experiment: p335
Performed by: Administrator | Machine: ABI | Report date: 1-10-2011

Sample name	Analysis status	Sample type	Gender	TP F/E	RP F/E	[DNA]	[DD]	[Primer%]
P335-A2 lot 0809 BMP-ID 05-HYI	ComparativeAnalysed	Sample	Female	41 / 41	10 / 10	OK	OK	OK 11%
P335-A2 lot 0809 BMP-ID 27-HYI	ComparativeAnalysed	Sample	Female	41 / 41	10 / 10	OK	OK	OK 10%
P335-A2 lot 0809 BMP-ID 28-HYI	ComparativeAnalysed	Sample	Female	41 / 41	10 / 10	OK	OK	OK 6%
P335-A2 lot 0809 BMP-ID 31-HYI	ComparativeAnalysed	Sample	Male	41 / 41	10 / 10	OK	OK	OK 6%
P335-A2 lot 0809 BMP-ID 32-HYI	ComparativeAnalysed	Sample	Male	41 / 41	10 / 10	OK	Warning	OK 8%
P335-A2 lot 0809 BMP-ID 43-HYI	ComparativeAnalysed	Sample	Male	41 / 41	10 / 10	OK	OK	OK 9%
P335-A2 lot 0809 BMP-ID 47-HYI	ComparativeAnalysed	Sample	Male	41 / 41	10 / 10	OK	OK	OK 7%
P335-A2 lot 0809 BMP-ID 55-HYI	ComparativeAnalysed	Positive	Male	41 / 41	10 / 10	OK	OK	OK 5%
P335-A2 lot 0809 BMP-ID 57-HYI	ComparativeAnalysed	Sample	Female	41 / 41	10 / 10	OK	OK	OK 8%
P335-A2 lot 0809 BMP-ID 74-HYI	ComparativeAnalysed	Sample	Male	41 / 41	10 / 10	OK	OK	OK 10%
P335-A2 lot 0809 BMP-ID 81-HYI	ComparativeAnalysed	Sample	Male	41 / 41	10 / 10	OK	OK	OK 9%

Figure 41 *Part of the sample quality control list of the experiment pdf report.*



Project: p335 | Experiment: p335
Performed by: Administrator | Machine: ABI | Report date: 1-10-2011

D [nt]	Gene-Exon	Chr.band	Hg18 loc.	Probe Nr.	[S] P335-A2 lot 080 9 BMP-ID 05-HYI-...	[S] P335-A2 lot 080 9 BMP-ID 27-HYI-...	[S] P335-A2 lot 080 9 BMP-ID 28-HYI-...	[S] P335-A2 lot 080 9 BMP-ID 31-HYI-...	[S] P335-A2 lot 080 9 BMP-ID 32-HYI-...	[S] P335-A2 lot 080 9 BMP-ID 43-HYI-...	[S] P335-A2 lot 080 9 BMP-ID 47-HYI-...	[P] P335-A2 lot 080 9 BMP-ID 55-HYI-...	[S] P335-A2 lot 080 9 BMP-ID 57-HYI-...	[S] P335-A2 lot 080 9 BMP-ID 74-HYI-...	[S] P...
454	EBF1-16	05q33.3	05-158057828	12527-13577	0.96	0.92	0.95	1.01	1.00	0.96	0.95	0.99	1.03	0.99	0.9
436	EBF1-14	05q33.3	05-158071810	13868-15386	0.98	0.95	0.94	0.99	0.94	0.98	1.02	1.00	1.00	0.96	1.0
364	EBF1-10	05q33.3	05-158137017	14059-15657	1.04	1.03	0.99	0.99	1.01	0.98	1.02	1.01	0.99	0.96	1.0
228	EBF1-1	05q33.3	05-158459187	12509-14269	0.93	0.97	0.91	0.96	0.92	1.01	***0.54****	0.97	0.94	0.98	0.9
269	IKZF1-1	07p12.2	07-050315024	13877-15918	1.03	***1.35****	1.09	***0.55****	1.11	0.99	0.97	0.96	1.03	1.03	1.0
208	IKZF1-2	07p12.2	07-050329206	14056-15654	0.97	1.19	1.01	***0.53****	1.00	0.97	0.92	0.99	0.97	0.95	1.0
177	IKZF1-3	07p12.2	07-050337794	12536-13586	0.98	1.21	0.84**	***0.50****	0.88	1.05	1.03	0.97	0.97	1.00	0.9
263	IKZF1-4	07p12.2	07-050411797	13873-15917	1.02	***1.37****	1.04	***0.55****	1.08	0.98	0.96	0.99	0.99	0.98	1.0
142	IKZF1-5	07p12.2	07-050417749	13872-15390	0.95	1.20	0.96	***0.53****	0.99	1.01	0.99	0.96	0.94	1.02	0.9
465	IKZF1-6	07p12.2	07-050422507	14061-15923	0.98	***1.39****	0.96	***0.53****	0.99	1.00	1.00	0.95	0.99	1.03	1.0
337	IKZF1-7	07p12.2	07-050426950	13869-15387	0.88	1.15	0.89	***0.53****	0.86	0.98	1.02	0.96	0.94	1.01	0.9
288	IKZF1-8	07p12.2	07-050436025	12549-15921	1.00	1.26	0.98	***0.50****	1.00	0.99	0.97	1.00	0.98	0.94	0.9
256	CDKN2A-5	09p21.3	09-021957822	10334-10871	***0.56****	0.89	0.89	0.93	0.91	***0.43****	0.94	***0.17****	***0.16****	0.93	0.9
251	CDKN2A-2a	09p21.3	09-021964958	10333-15916	***0.66****	1.04	0.93	0.92	1.03	***0.43****	1.12	***0.20****	***0.11****	0.87	0.8
238	CDKN2B-2	09p21.3	09-021995813	10337-15914	***0.13****	0.98	0.83**	0.94	1.00	***0.40****	1.00	***0.18****	***0.15****	0.95	0.9
373	PAX5-10	09p13.2	09-036830475	12521-13571	1.17	0.93	1.10	0.97	0.96	***0.68****	0.92	0.98	***0.55****	0.90	1.0
275	PAX5-8	09p13.2	09-036872066	14057-15919	1.00	1.03	0.91	1.02	0.91	0.98	0.96	***0.87****	***0.56****	0.97	0.9
202	PAX5-6	09p13.2	09-036956622	14055-15653	0.96	0.83**	0.89	0.96	0.86	0.95	***0.58****	0.97	0.89	1.06	0.9
282	PAX5-5	09p13.2	09-036992699	13870-15920	1.05	0.95	1.01	1.03	0.97	1.03	***0.58****	***2.85****	1.08	0.99	0.9
483	PAX5-2	09p13.2	09-037010669	13876-15924	1.11	0.89	1.00	1.07	1.10	0.99	***0.53****	***2.64****	1.02	1.05	1.0
154	PAX5-1	09p13.2	09-037024272	12501-13551	1.03	0.98	1.03	1.01	1.07	1.02	0.96	0.99	1.02	0.95	1.0
301	ETV6-1	12p13.2	12-011694211	14058-15656	***0.52****	0.98	1.03	1.06	1.03	1.09	1.09	1.05	1.07	1.01	1.0
382	ETV6-1	12p13.2	12-011694328	14060-15658	***0.58****	0.92	1.04	1.02	0.97	1.05	1.01	1.01	1.04	1.01	1.0
391	ETV6-2	12p13.2	12-011796704	13875-15393	***0.62****	0.97	1.02	1.01	1.04	0.79**	***0.56****	1.02	1.05	1.03	1.0
246	ETV6-3	12p13.2	12-011883341	13874-15915	***0.56****	0.98	1.01	0.98	1.04	1.00	***0.54****	1.04	1.06	0.97	0.9
474	ETV6-5	12p13.2	12-011913651	13871-15389	***0.57****	1.03	1.10	1.04	1.13	1.07	***0.64****	1.05	1.10	0.94	1.0
196	ETV6-8	12p13.2	12-011935363	14054-15652	1.05	0.93	0.97	0.99	1.00	0.98	***0.59****	1.01	0.99	1.06	1.0
409	BTG1 AREA-up	12q21.33	12-090903053	14117-15718	1.07	1.00	1.00	0.96	0.96	1.05	0.98	***0.41****	1.02	1.01	1.0
160	BTG1 AREA-up	12q21.33	12-090905485	14116-15717	0.94	1.07	1.03	0.98	1.02	0.98	1.00	***0.41****	1.00	0.95	0.9
328	BTG1-2	12q21.33	12-091061700	12553-13603	1.06	1.13	1.04	0.99	1.10	1.03	1.00	***0.42****	1.01	0.96	0.9
232	BTG1-1	12q21.33	12-091063286	12542-15913	0.92	0.96	0.91	0.94	0.94	1.04	0.98	0.97	0.94	1.02	0.9

Figure 42 *Part of the experiment pdf report ratio overview.*

8.3.3 Comparative analysis experiment explorer statistical overview grid

The third tab displays the same data as described at the 8.3.1 but now in grid format (figure 43).

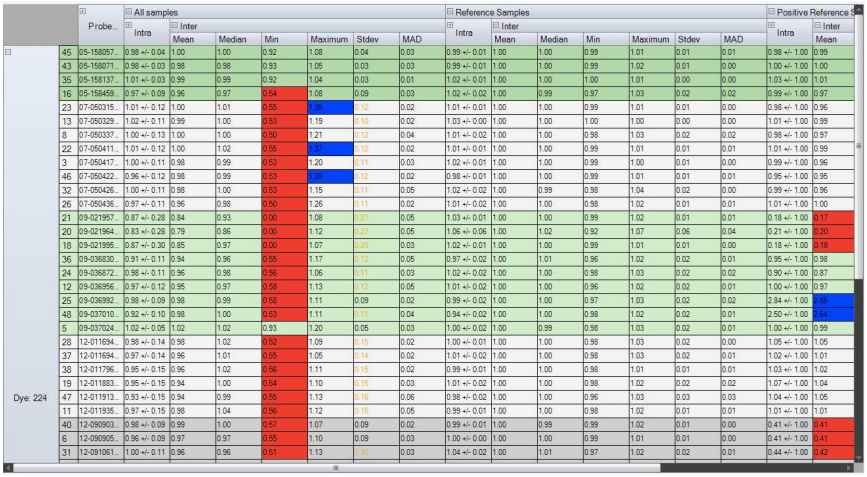


Figure 43 *Experiment explorer statistical overview grid.*

The statistical overview grid displays data in a similar fashion as the previously described heat map grid, however instead of displaying the samples and their data this grid display the calculated statistical value over each sample type. For each probe the following statistical value are calculated over all sample of the same sample type: average, median, minimum, maximum, standard deviation and MAD (median of absolute deviations). By using right mouse click menu, region coloring may be changed or the grid may be exported in a similar way as described for the heat map grid. The right mouse click context menu contains the same options as earlier described for the ratio overview grid.

8.4 About the comparative sample results explorer

The comparative sample results explorer may provide a more detailed view to the results of each sample separately as opposed to viewing the reporting option of the experiment explorer. To open the experiment explorer: right mouse click on the grid showing the quality scores and select from the right click menu “Open sample results”. Alternatively the sample results explorer

may be opened through the experiment explorer as described in 8.3.2. The comparative analysis sample explorer has three tabs allowing getting a comprehended view of the results of the selected sample and the statistical significance of that result within the experiment.

8.4.1 Comparative analysis sample explorer statistical sample chart

The first tab opens shows a sample chart displaying the ratios results of the last normalization step (figure 44). On the Y-axis the probe ratios are displayed of the sample that was selected in the left list box. You may switch samples by either using the cursor keys or by selecting a sample from the list use a mouse click. After each sample of the type “reference sample” you will find the tag “[r]” and sample of the type “positive reference” will have an added tag “[p]”.

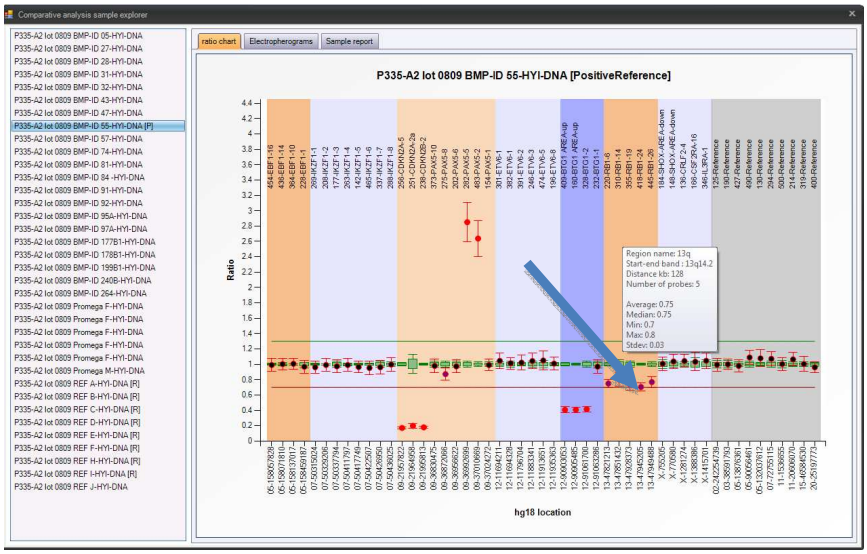


Figure 44 *Sample explorer ratio chart*

Each black, red or purple circular marker points indicate the result of a single probe in the selected sample. On default the X-axis loads with the hg18 track map view locations and the labels display a “probe design length probe gene name – probe gene exon number” notation. The found whiskers at each probe marker ratio indicate the estimated 95% confidence range for that signal. These confidence ranges are estimated by combining the found discrepancies of the estimated dosage quotients by the used reference

probes and/or reference samples. The estimated variability of each probe in the used reference collection may thus provide information if that probe was found to be reproducible in the performed experiment and the variability found over the used reference probes may indicate if the quality of the normalization was adequate. For more information about these calculations please see published articles about the methodology behind Cofaflyser.NET (J. Coffa, 2011; J. Coffa 2008).

Results of all samples of the same sample type for each probe are be displayed by a box plot, differently from the earlier discussed box plot this one displays the estimated 68.1% (1 standard deviation) confidence range by the box and the 95% confidence range by the outer whiskers (2 standard deviations). The statistics over the reference sample collection are loaded on default by a green box, test sample population by a blue box and the positive sample population by a yellow box. A found single probe result thus has a higher probability to be different from the reference population if the estimated 95% confidence range of that signal does not overlap with the outer whiskers of the green box or 95% confidence range of the reference sample population. Single sample probe results that fall within the 95% confidence range of the reference sample population will be displayed as by black round markers (figure 45a); if the results fall outside of the 95% confidence range but are still between the set arbitrary borders, they will be displayed by purple circular markers (figure 45b), and in case they also fall outside the arbitrary borders they will be displayed as red round markers (figure 45c & 45d). Finally we may find results that fall within the 95% confidence range of the reference sample population but fall outside of the set arbitrary border. Such contrary results are marked by a salmon colored circular marker and are also called ambiguous (figure 45e & 45f). Please note that these different result stages accord to the population comparison values as described at 8.3.2.

The displayed regions listen to the same functionality as described in 8.3.1 at the comparative analysis experiment explorer statistical overview chart. The tool tip controls display the basic statistics for all the probes that fall within that region based on their final estimated ratios. Chromosomal aberrations often-span larger regions (M. Hermsen, 2002), which allow probes targeted to that region to cluster together by sorting. This data may aid in determination if all signals of the probes that fall in one region are either or decreases as opposed to a certain population. In figure 49 for instance shows a case where all signals of the probes targeted to 13q14.2 are decreased with 25%. The median ratio of that region was 0.75 with a standard deviation of 0.03 indicating that this sample probably contains a mixed cell population where 50% of the cells harbor a heterozygous deletion for 13q14.2 while the other 50% of the cells originate from cells that are diploid for 13q14.2.

The right click menu enables you to customize the chart. The exact same options can be found in this menu as earlier described at 8.3.1 for the comparative analysis experiment explorer statistical overview chart.

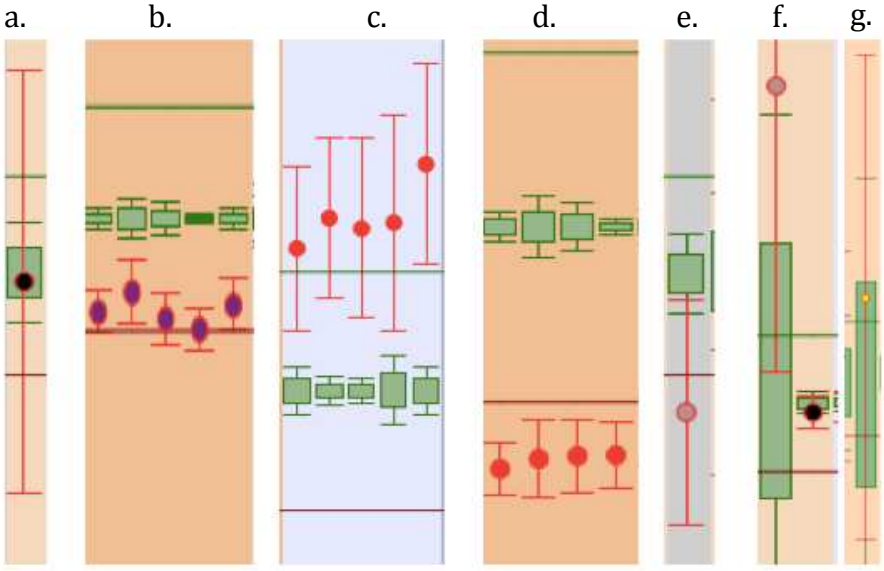


Figure 45 *Different probe ratio results stages versus the reference sample populations. Result "a" displays a result that was found to be equal to the reference sample population. Result "b" was found to be significantly different from the reference population but did not fall outside the arbitrary borders. Such cases are often seen when samples have mosaic cell populations. Result "c" is significantly increased as opposed to the reference population and the result is also higher than the set arbitrary borders. Result "d" is significantly decreased as opposed to the reference population and the result is also lower than the set arbitrary borders. Results "e" and "f" are both half-ambiguous. Result "e" is lower than the set arbitrary border but the reference probes for this sample were variable, the result was therefore found to be different from only 68% of the reference population and not 95%. Results "f" on the other hand shows a very wide 95% confidence range for that probe in the reference sample collection, indicating low reproducibility for that probe in the experiment. Again this result was found to be different from only 68% of the reference population. Results "g" is ambiguous, the ratio of the signal is higher than the set arbitrary border but both the reference probes and reference samples for this sample were variable resulting in a large standard deviation for sample probe result making the result inconclusive.*

8.4.2 Comparative analysis sample explorer electropherogram viewer

The second tab allows you to explore the original electropherograms better. Confirmation of found results is often not only desirable but also imperative in order to get an indisputable assessment. The displayed electropherogram descent directly from the baseline corrected original data stream created by your capillary electrophoresis device (figure 46). Even though a line chart may be visible the original data stream consists of separate time points or data points. The sample electropherogram tab on default present the data point on the x-axis and the relative fluorescent units on the y-axis. To make data interpretation easier the design probe length are displayed underneath the x-axis at the data point level of the detected peak top of that probe. A peak that was related to a probe will furthermore have a circular marker at the detected peak top data point to relative fluorescent units.

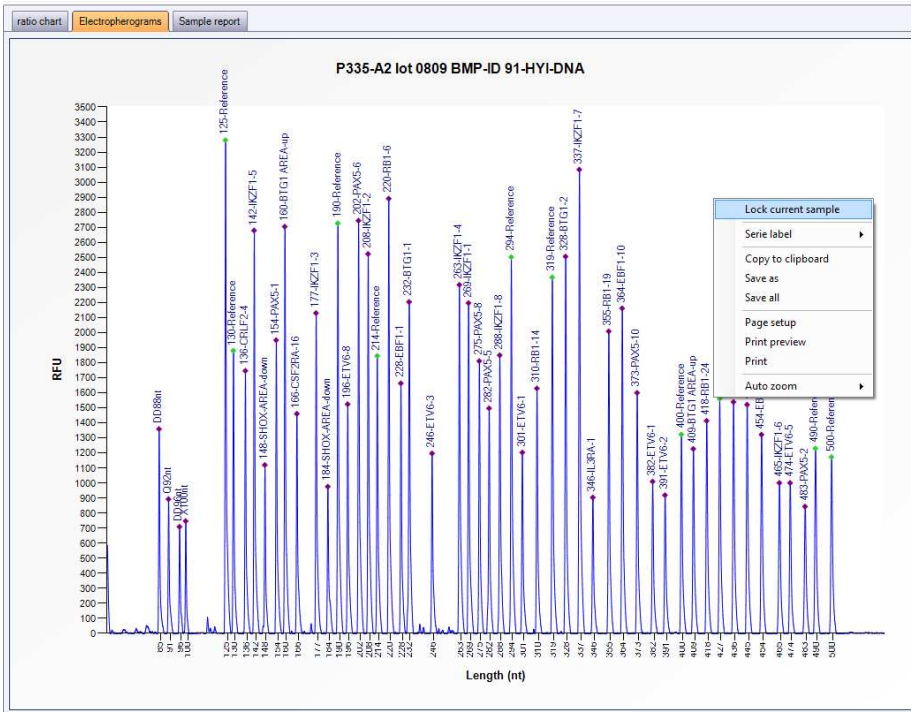


Figure 46 *Comparative analysis sample explorer electropherogram tab.*

By hovering above this marker you may view different information about this detection peak including probe target information and probe ratio at the different stages of analysis (figure 47). In the right mouse click menu you may find options allowing you to export, print, save, zoom and adjust the

labels of the chart. The option “Lock current sample” in the right click menu will split the chart area in two. The upper part of the chart area will then show the results of the sample that is displayed at the moment of locking, while the lower part listens to the original functionality (figure 47). Automatic zooming allows 3 levels of zoom, this being: show all detected peaks, show all recognized peaks and show all peaks recognized as MLPA test probes. You may furthermore use manual zooming by clicking anywhere in the chart and dragging the mouse over the area you wish to zoom into. Zooming in double view modus always automatically perform a zoom on the both parts of the chart. It should be noted that due to differences in separation speed between different channels, peaks might appear to be a slightly different positions. At full automatic zoom methods there will be corrected for these differences.

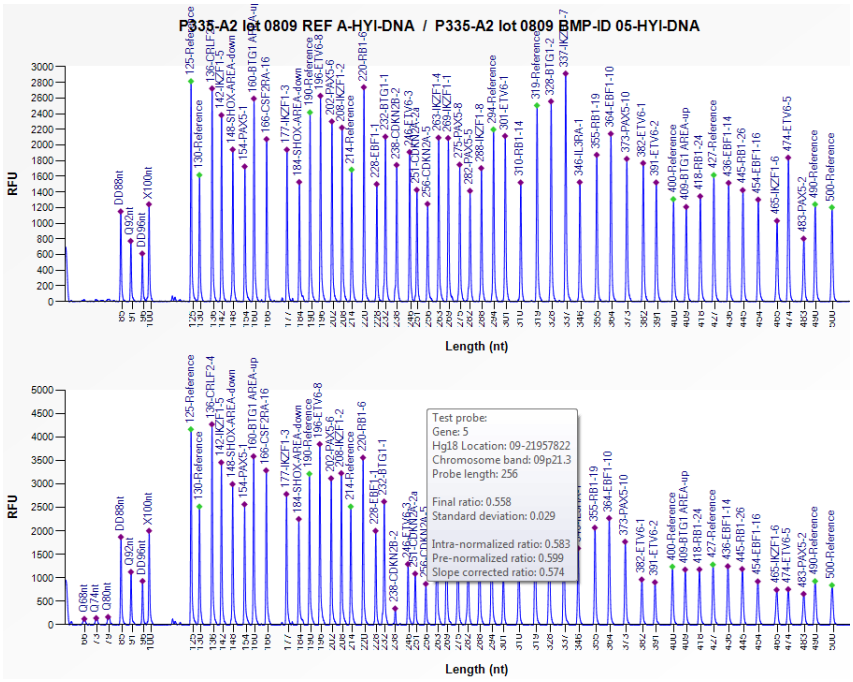


Figure 47 *Comparative analysis sample explorer electropherogram double sample view; the displayed tool tip box shows a peak signal found to have a reduction of 50% as opposed to the reference population. Top part of the chart area displays the electropherogram of a reference sample while the bottom shows that of a tumor sample both tested with P335 MLPA mix.*

8.4.3 Comparative analysis sample explorer report viewer

The third and final tab can be used for reporting services. The displayed grid shows information concerning the target sequence of each probe and also all relevant information of the peak signal that was related to that probe (figure 48). While most displayed fields are already explained at 8.3.2 there are two extra columns which were not discussed yet, the RSQ (reference sample quality) column and the RPQ (reference probe quality) column. The RSQ accounts for the part of the standard deviation of each probe that is calculated over the ratios when applying multiple reference samples. The RPQ account for the part by the usage of multiple reference probes. The final standard deviation is estimated by combining these two 2 factors, as explained at 8.4.1. By using the right mouse menu this grid may be exported to a file in *.*csv, *.*HTML, *.*XML document or *.*XML spreadsheet format. More important by using the right mouse click sample pdf reports may be generated. Coffalyser.NET allows the generation of two types of pdf sample reports. A single page report where all data of the three sample explorer tabs are put together in landscape modus (figure 49) and a two-page report, which also contains extended information.

ratio chart | Electropherograms | Sample report

	Name	Chr band	hg18	Length	Order	Signal	Intra ratio	Pre ratio	Ratio wo iteration	Final ratio	[RSQ]	[RPQ]	Stdev	RefPop	SamPop
45	EBF1-16	05q33.3	05-158...	454	001	923	0.96	0.91	0.98	0.97	0.01	0.07	0.05	=	=
43	EBF1-14	05q33.3	05-158...	436	002	1246	0.99	0.94	1	0.99	0.01	0.06	0.04	=	=
35	EBF1-10	05q33.3	05-158...	364	003	2271	1.07	1.04	1.06	1.03	0	0.06	0.04	=	=
16	EBF1-1	05q33.3	05-158...	228	004	2000	0.96	1	0.95	0.92	0.02	0.06	0.05	=	=
23	IKZF1-1	07p12.2	07-050...	269	005	2514	1.04	1.06	1.02	1.01	0.02	0.06	0.05	=	=
13	IKZF1-2	07p12.2	07-050...	208	006	3229	1.01	1.08	1	0.98	0	0.08	0.06	=	=
8	IKZF1-3	07p12.2	07-050...	177	007	2784	1.01	1.09	1	0.99	0.02	0.07	0.05	=	=
22	IKZF1-4	07p12.2	07-050...	263	008	2632	1.04	1.06	1.02	1.01	0.01	0.07	0.05	=	=
3	IKZF1-5	07p12.2	07-050...	142	009	3458	0.99	1.09	0.99	0.96	0.02	0.06	0.04	=	=
46	IKZF1-6	07p12.2	07-050...	465	010	744	0.98	0.93	1	0.99	0.02	0.06	0.05	=	=
32	IKZF1-7	07p12.2	07-050...	337	011	2719	0.91	0.9	0.9	0.88	0.03	0.05	0.04	=	=
26	IKZF1-8	07p12.2	07-050...	288	012	2019	1.02	1.03	1.01	1	0.01	0.07	0.05	=	=
21	CDKN2...	09p21.3	09-021...	256	013	871	0.58	0.6	0.57	0.56	0.01	0.04	0.03	<<	<
20	CDKN2...	09p21.3	09-021...	251	014	1088	0.71	0.74	0.71	0.69	0.05	0.04	0.06	<<	?
18	CDKN2...	09p21.3	09-021...	238	015	346	0.13	0.14	0.13	0.13	0	0.01	0.01	<<	<<
36	PAX5-10	09p13.2	09-036...	373	016	1769	1.17	1.14	1.16	1.16	0.03	0.08	0.07	=	=
24	PAX5-8	09p13.2	09-036...	275	017	1914	1.04	1.06	1.03	1	0.02	0.07	0.05	=	=
12	PAX5-6	09p13.2	09-036...	202	018	3119	0.99	1.05	0.98	0.95	0.02	0.06	0.05	=	=
25	PAX5-5	09p13.2	09-036...	282	019	1561	1.07	1.08	1.05	1.07	0.03	0.07	0.06	=	=
48	PAX5-2	09p13.2	09-037...	483	020	657	1.07	1.02	1.12	1.12	0.03	0.08	0.06	=	=
5	PAX5-1	09p13.2	09-037...	154	021	2565	1.04	1.14	1.04	1.04	0.02	0.06	0.05	=	=
28	ETV6-1	12p13.2	12-011...	301	022	1121	0.53	0.53	0.52	0.52	0.01	0.04	0.03	<<	<<
37	ETV6-1	12p13.2	12-011...	382	023	965	0.59	0.57	0.59	0.58	0.01	0.04	0.03	<<	<<
38	ETV6-2	12p13.2	12-011...	391	024	905	0.62	0.6	0.62	0.61	0.01	0.04	0.03	<<	<<
19	ETV6-3	12p13.2	12-011...	246	025	1293	0.56	0.58	0.55	0.54	0.01	0.03	0.03	<<	<<

Figure 48 *Comparative analysis sample report grid.*

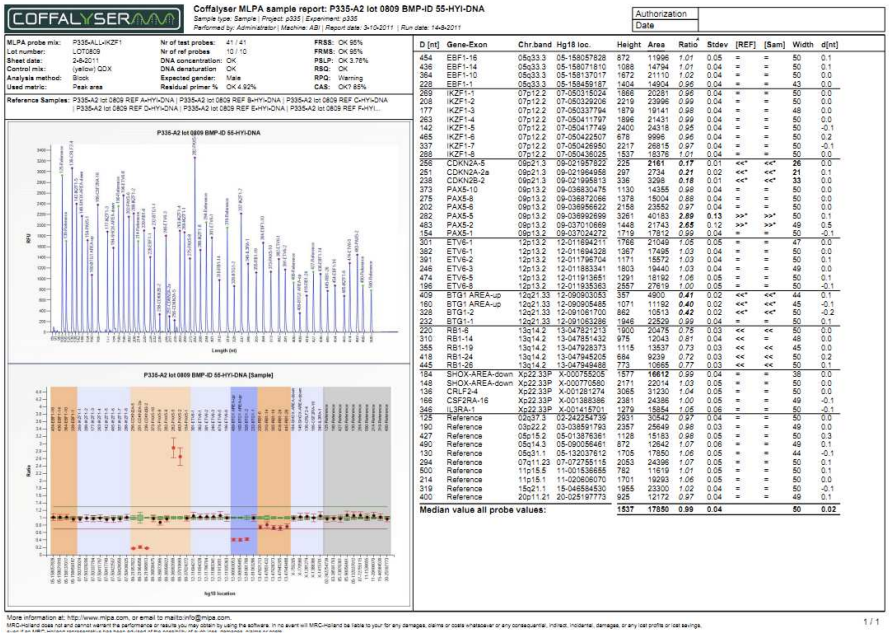
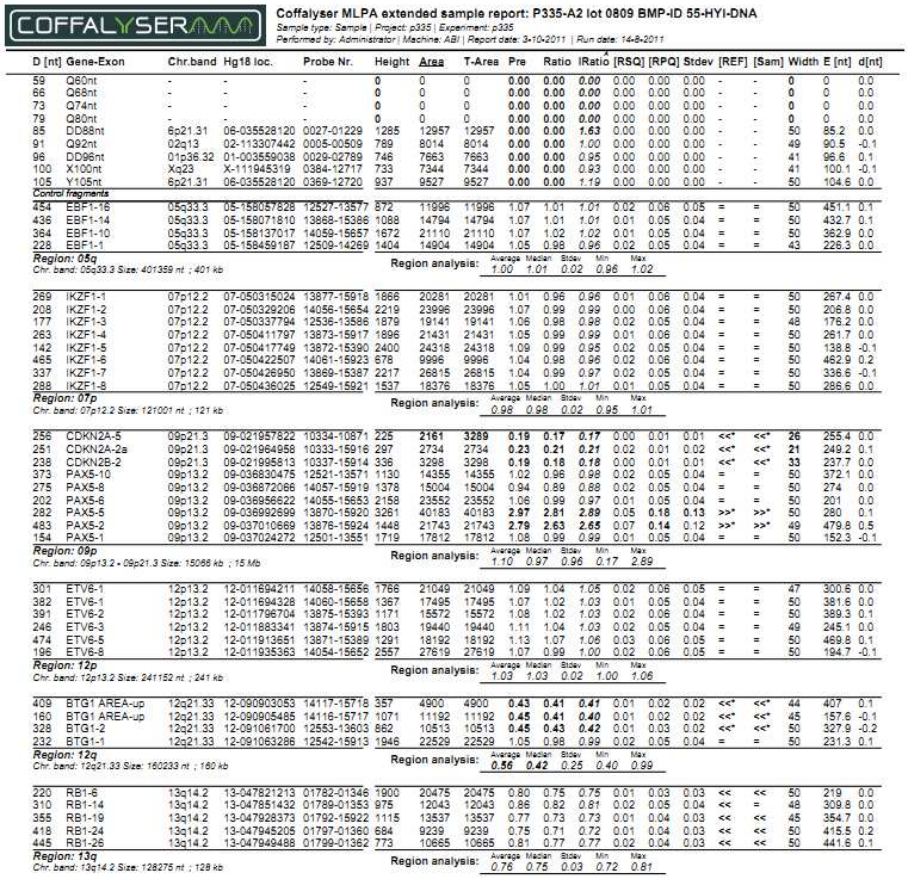

COFFALYSER
Coffalyser MLPA sample report: P335-A2 lot 0809 BMP-ID 55-HYI-DNA
Authorization
Date
P335-A2 lot 0809 BMP-ID 55-HYI-DNA
P335-A2 lot 0809 BMP-ID 55-HYI-DNA [Sample]
More information at: http://www.mlpa.com, or email to mailto:info@mlpa.com.
1 / 1

Figure 49 *Single page sample pdf report.*

The dual page contains all relevant quality control information on the first page together with a larger sample chart and electropherogram. The second page (figure 50) contains a report of all probes and their target information. Next to this the peak height, peak area, total peak area in the probe bin, population normalized ratio, slope corrected ratio, final ratio, reference sample quality standard deviation, reference probe standard deviation, final standard deviation, distribution comparison values, peak width, expected peak length and delta to that expected length are also added in the report. In case any of the columns contain values that were found to differ from the rest they will become bold. Note that the expected lengths are the lengths of the peak that were used as the center for data filtering. These values are commonly based on the entire data set and peaks are not expected to differ much from their expected length (<0.5 nt).



COFFALYSER

Coffalyser MLPA extended sample report: P335-A2 lot 0809 BMP-ID 55-HYI-DNA

Sample type: Sample | Project: p335 | Experiment: p335

Performed by: Administrator | Machine: ABI | Report date: 3-10-2011 | Run date: 14-8-2011

D [nt]	Gene-Exon	Chr.band	Hg18 loc.	Probe Nr.	Height	Area	T-Area	Pre	Ratio	▲Ratio	[RSQ]	[RPQ]	Stdev	[REF]	[Sam]	Width	E [nt]	d[nt]
59	Q60nt	-	-	-	0	0	0	0.00	0.00	0.00	0.00	0.00	0.00	-	-	0	0	0.0
66	Q68nt	-	-	-	0	0	0	0.00	0.00	0.00	0.00	0.00	0.00	-	-	0	0	0.0
73	Q74nt	-	-	-	0	0	0	0.00	0.00	0.00	0.00	0.00	0.00	-	-	0	0	0.0
79	Q80nt	-	-	-	0	0	0	0.00	0.00	0.00	0.00	0.00	0.00	-	-	0	0	0.0
85	DD88nt	6p21.31	06-035528120	0027-01229	1285	12957	12957	0.00	0.00	1.63	0.00	0.00	0.00	-	-	50	85.2	0.0
91	Q92nt	02q13	02-113307442	0005-00509	789	8014	8014	0.00	0.00	1.00	0.00	0.00	0.00	-	-	49	90.5	-0.1
96	DD96nt	01p36.32	01-003559038	0029-02789	746	7663	7663	0.00	0.00	0.95	0.00	0.00	0.00	-	-	41	96.6	0.1
100	X100nt	Xq23	X-111945319	0384-12717	733	7344	7344	0.00	0.00	0.93	0.00	0.00	0.00	-	-	41	100.1	-0.1
105	Y105nt	6p21.31	06-035528120	0369-12720	937	9527	9527	0.00	0.00	1.19	0.00	0.00	0.00	-	-	50	104.6	0.0
Control fragments																		
454	EBF1-16	05q33.3	05-158057828	12527-13577	872	11996	11996	1.07	1.01	1.01	0.02	0.06	0.05	=	=	50	451.1	0.1
436	EBF1-14	05q33.3	05-158071810	13868-15386	1088	14794	14794	1.07	1.01	1.01	0.01	0.05	0.04	=	=	50	432.7	0.1
384	EBF1-10	05q33.3	05-158137017	14059-15657	1672	21110	21110	1.07	1.02	1.02	0.01	0.05	0.04	=	=	50	382.9	0.0
228	EBF1-1	05q33.3	05-158459187	12509-14269	1404	14904	14904	1.05	0.98	0.96	0.02	0.05	0.04	=	=	43	226.3	0.0
Region: 05q								Region analysis:	Average	Median	Stdev	Min	Max					
Chr. band: 05q33.3 Size: 401359 nt ; 401 kb									*1.00*	*1.01*	*0.02*	*0.96*	*1.02*					
269	IKZF1-1	07p12.2	07-050315024	13877-15918	1866	20281	20281	1.01	0.96	0.96	0.01	0.06	0.04	=	=	50	267.4	0.0
208	IKZF1-2	07p12.2	07-050329206	14056-15654	2219	23996	23996	1.07	0.99	0.99	0.00	0.06	0.04	=	=	50	206.8	0.0
177	IKZF1-3	07p12.2	07-050337794	12536-13586	1879	19141	19141	1.06	0.98	0.98	0.02	0.05	0.04	=	=	48	176.2	0.0
263	IKZF1-4	07p12.2	07-050411797	13873-15917	1896	21431	21431	1.05	0.99	0.99	0.01	0.06	0.04	=	=	50	261.7	0.0
142	IKZF1-5	07p12.2	07-050417749	13872-15390	2400	24318	24318	1.09	0.99	0.95	0.02	0.05	0.04	=	=	50	138.8	-0.1
465	IKZF1-6	07p12.2	07-050422507	14061-15923	678	9996	9996	1.04	0.98	0.96	0.02	0.06	0.04	=	=	50	462.9	0.2
337	IKZF1-7	07p12.2	07-050426950	13869-15387	2217	26815	26815	1.04	0.99	0.97	0.02	0.05	0.04	=	=	50	336.6	-0.1
288	IKZF1-8	07p12.2	07-050436025	12549-15921	1537	18376	18376	1.05	1.00	1.01	0.01	0.05	0.04	=	=	50	286.6	0.0
Region: 07p								Region analysis:	Average	Median	Stdev	Min	Max					
Chr. band: 07p12.2 Size: 121001 nt ; 121 kb									*0.98*	*0.98*	*0.02*	*0.95*	*1.01*					
256	CDKN2A-5	09p21.3	09-021957822	10334-10871	225	2161	3289	0.19	0.17	0.17	0.00	0.01	0.01	<<*	<<*	26	255.4	0.0
251	CDKN2A-2a	09p21.3	09-021964958	10333-15916	297	2734	2734	0.23	0.21	0.21	0.02	0.01	0.02	<<*	<<*	21	249.2	0.1
238	CDKN2B-2	09p21.3	09-021995813	10337-15914	336	3298	3298	0.19	0.18	0.18	0.00	0.01	0.01	<<*	<<*	33	237.7	0.0
373	PAX5-10	09p13.2	09-036830475	12521-13571	1130	14355	14355	1.02	0.98	0.98	0.02	0.05	0.04	=	=	50	372.1	0.0
275	PAX5-8	09p13.2	09-036872066	14057-15919	1378	15004	15004	0.94	0.89	0.89	0.02	0.05	0.04	=	=	50	274	0.0
202	PAX5-6	09p13.2	09-036956622	14055-15653	2158	23552	23552	1.06	0.99	0.97	0.01	0.05	0.04	=	=	50	201	0.0
282	PAX5-5	09p13.2	09-036992699	13870-15920	3261	40183	40183	2.97	2.81	2.89	0.05	0.18	0.13	>>*	>>*	50	280	0.1
483	PAX5-2	09p13.2	09-037010669	13876-15924	1448	21743	21743	2.79	2.63	2.65	0.07	0.14	0.12	>>*	>>*	49	479.8	0.5
154	PAX5-1	09p13.2	09-037024272	12501-13551	1719	17812	17812	1.08	0.99	0.99	0.01	0.05	0.04	=	=	50	152.3	-0.1
Region: 09p								Region analysis:	Average	Median	Stdev	Min	Max					
Chr. band: 09p13.2 - 09p21.3 Size: 15066 kb ; 15 Mb									*1.10*	*0.97*	*0.96*	*0.17*	*2.89*					
301	ETV6-1	12p13.2	12-011694211	14058-15656	1766	21049	21049	1.09	1.04	1.05	0.02	0.06	0.05	=	=	47	300.6	0.0
382	ETV6-1	12p13.2	12-011694328	14060-15658	1367	17495	17495	1.07	1.02	1.03	0.01	0.05	0.04	=	=	50	381.6	0.0
391	ETV6-2	12p13.2	12-011796704	13875-15393	1171	15572	15572	1.08	1.02	1.03	0.02	0.06	0.04	=	=	50	389.3	0.1
246	ETV6-3	12p13.2	12-011883341	13874-15915	1803	19440	19440	1.11	1.04	1.03	0.02	0.05	0.04	=	=	49	245.1	0.0
474	ETV6-5	12p13.2	12-011913651	13871-15389	1291	18192	18192	1.13	1.07	1.06	0.03	0.06	0.05	=	=	50	469.8	0.1
196	ETV6-8	12p13.2	12-011935363	14054-15652	2557	27619	27619	1.07	0.99	1.00	0.02	0.06	0.05	=	=	50	194.7	-0.1
Region: 12p								Region analysis:	Average	Median	Stdev	Min	Max					
Chr. band: 12p13.2 Size: 241152 nt ; 241 kb									*1.03*	*1.03*	*0.02*	*1.00*	*1.06*					
409	BTG1 AREA-up	12q21.33	12-090903053	14117-15718	357	4900	4900	0.43	0.41	0.41	0.01	0.02	0.02	<<*	<<*	44	407	0.1
160	BTG1 AREA-up	12q21.33	12-090905485	14116-15717	1071	11192	11192	0.45	0.41	0.40	0.01	0.02	0.02	<<*	<<*	45	157.6	-0.1
328	BTG1-2	12q21.33	12-091061700	12553-13603	862	10513	10513	0.45	0.43	0.42	0.01	0.03	0.02	<<*	<<*	50	327.9	-0.2
232	BTG1-1	12q21.33	12-091063286	12542-15913	1948	22529	22529	1.05	0.98	0.99	0.02	0.05	0.04	=	=	50	231.3	0.1
Region: 12q								Region analysis:	Average	Median	Stdev	Min	Max					
Chr. band: 12q21.33 Size: 160233 nt ; 160 kb									*0.56*	*0.42*	*0.25*	*0.40*	*0.99*					
220	RB1-6	13q14.2	13-047821213	01782-01346	1900	20475	20475	0.80	0.75	0.75	0.01	0.03	0.03	<<	<<	50	219	0.0
310	RB1-14	13q14.2	13-047851432	01789-01353	975	12043	12043	0.86	0.82	0.81	0.02	0.05	0.04	<<	=	48	309.8	0.0
355	RB1-19	13q14.2	13-047928373	01792-15922	1115	13537	13537	0.77	0.73	0.73	0.01	0.04	0.03	<<	<<	45	354.7	0.0
418	RB1-24	13q14.2	13-047945205	01797-01360	684	9239	9239	0.75	0.71	0.72	0.01	0.04	0.03	<<	<<	50	415.5	0.2
445	RB1-26	13q14.2	13-047949488	01799-01362	773	10665	10665	0.81	0.77	0.77	0.02	0.04	0.03	<<	<<	50	441.6	0.1
Region: 13q								Region analysis:	Average	Median	Stdev	Min	Max					
Chr. band: 13q14.2 Size: 128275 nt ; 128 kb									*0.76*	*0.75*	*0.03*	*0.72*	*0.81*					

Figure 50 *Dual page sample extended pdf report page 2.*

9. Methylation specific MLPA analysis

9.1 Introduction to MS-MLPA analysis

MS-MLPA analysis can be applied as an extension on the normal copy number DNA-MLPA analysis. In Coffalyser.NET copy number and methylations status analysis always occurs in a single analysis. Results of copy number and methylations status are then displayed together making data interpretation easier. Interpreting MS-MLPA data with the copy number status is crucial since only relative methylation percentages of target sequences are calculated. Without copy number information these percentages would be very difficult to interpret. During a DNA/MS-MLPA analysis, the normal DNA-MLPA analysis is initially performed, normalizing all samples of the type “sample” and “positive reference” against all available samples of the type “reference sample”. After the calculation of all distribution statistics the MS-MLPA analysis will follow automatically. Here, each sample of the type “sample” is matched against available digested samples by using a Smith&Waterman algorithm on the sample name. To ensure that that this matching is successful is it recommended giving the cut and uncut samples equal names in the capillary sample sheets. Samples that are for instance named “Sample1-Undig” and “Sample1-dig” will ensure correct matching. After each sample is matched, the methylation status normalization will follow normalizing the data of the digested samples directly against their undigested counter parts. During this normalization only a single reference sample exists for each digested sample (the undigested counterpart).
The reproducibility of each probe in the experiment is therefore derived from the DNA-MLPA analysis. We thus assume that the reproducibility over the reference samples of each probe as found in the DNA-MLPA analysis can be applied to the reproducibility of the probes in both the copy number as the methylation status analysis. The reproducibility as determined over the reference probes is determined during the MS-MLPA normalization.
It should also be noted that Coffalyser.NET allows a separate reference probe selection for DNA-MLPA normalization and MS-MLPA normalization. For more information about how to change the reference probes used for the different normalization, please see chapter 3 about the MRC-MLPA sheet manager.

9.2 Setting up the MS-MLPA analysis

To perform a MS-MLPA analysis first set the experiment type to “DNA/MS-MLPA” as explained earlier at 6.3. Next import all your samples as you would normally do, however when settings the sample types set every digested sample to the sample type “digested”, as explained in chapter 6.4. Perform the fragment analysis and explore the fragment data as described earlier in chapter 7 and 8. Please note during data exploration that digested target sequences are expected to be absent and that these sample runs are therefore expected to have a lower normal number of peaks as compared to their undigested counterparts. Following the fragment analysis you need to match the digested and undigested samples in the comparative analysis settings screen. Right click to open the context menu and select “digested samples (for MS-MLPA)” and then select “match samples automatically” (figure 51). This will enable a matching algorithm based on the Smith and Waterman method adapted for sample names. Each undigested sample in the first column will be matched against a digested sample in the collection, which will afterwards appear in the column “digested”. Because matching may not always be 100% successful users may adapt the matched sample by double clicking on any of the cells in the column “digested” and change it into the corrected sample. Please note that each undigested sample can only be matched against one unique digested sample. After making all the correct matches click on “start comparative analysis”. All the settings can be made exactly as described earlier for the “DNA-MLPA” at chapter 8.1. Methylation specific normalization occurs always in the same way and the methodology cannot be adapted, the available settings thus only influence the analysis of the DNA-MLPA normalization. This method normalized each target test probe of each test sample directly against its undigested counterpart by making use of the set reference probes. This method does not require any slope correction since the sample is the same on both side of the equation and a difference in sloping between the two is not expected.

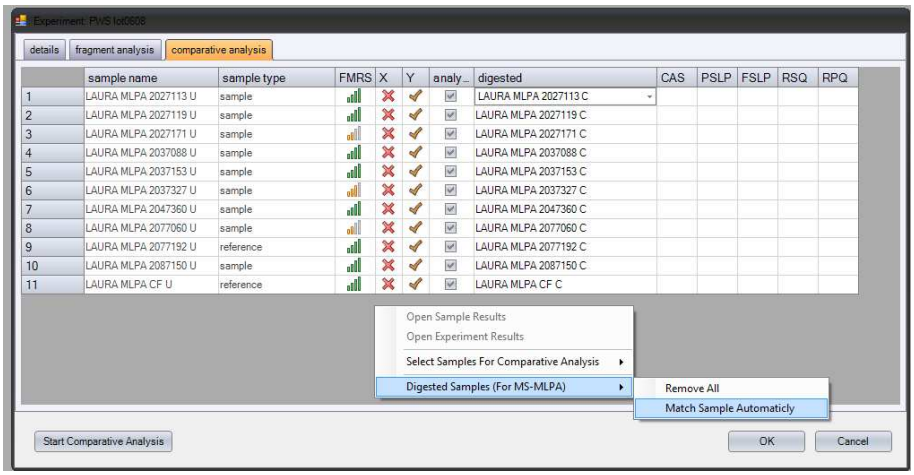

Experiment: PWS lot0608
details
fragment analysis
comparative analysis
sample name
sample type
FMRS
X
Y
analy...
digested
CAS
PSLP
FSLP
RSQ
RPQ
1 LAURA MLPA 2027113 U sample LAURA MLPA 2027113 C
2 LAURA MLPA 2027119 U sample LAURA MLPA 2027119 C
3 LAURA MLPA 2027171 U sample LAURA MLPA 2027171 C
4 LAURA MLPA 2037088 U sample LAURA MLPA 2037088 C
5 LAURA MLPA 2037153 U sample LAURA MLPA 2037153 C
6 LAURA MLPA 2037327 U sample LAURA MLPA 2037327 C
7 LAURA MLPA 2047360 U sample LAURA MLPA 2047360 C
8 LAURA MLPA 2077060 U sample LAURA MLPA 2077060 C
9 LAURA MLPA 2077192 U reference LAURA MLPA 2077192 C
10 LAURA MLPA 2087150 U sample LAURA MLPA 2087150 C
11 LAURA MLPA CF U reference LAURA MLPA CF C
Open Sample Results
Open Experiment Results
Select Samples For Comparative Analysis
Digested Samples (For MS-MLPA)
Remove All
Match Sample Automaticly
Start Comparative Analysis
OK
Cancel

Figure 51 *Comparative analysis settings screen in DNA/MS-MLPA modus.*

9.3 Comparative analysis experiment explorer with MS-Data

After the analysis is finished you can reopen the comparative analysis experiment explorer as earlier described in chapter 8.3. When you select distribution type, you will find extra distribution for each sample type separately for the DNA and MS-MLPA analysis. You may for instance display the results of the “reference samples MS” you can easily evaluate the reproducibility of each probe in that experiment for the methylation status, assuming the selected reference samples were genetically equal and the reference samples were properly dispersed through the experiment. In the heat map grid, each digested sample MS-result will be loaded directly next to its undigested sample DNA-result. In figure 52 you may for instance view the DNA/MS-MLPA results of the ME028 Prader Willi mix. In the left column you may find the reference samples which are normal ratio 1 for the DNA-MLPA results, while the SNRPN have a normal methylation status of 50%, in these samples displayed by a red cell ratio 0.5. Other probes also known as digestion control probes will not have a signal at all. The tab of the comparative analysis experiment explorer containing the statistical overview grid will also automatically be extended with a separate level for each sample type for all methylation results.

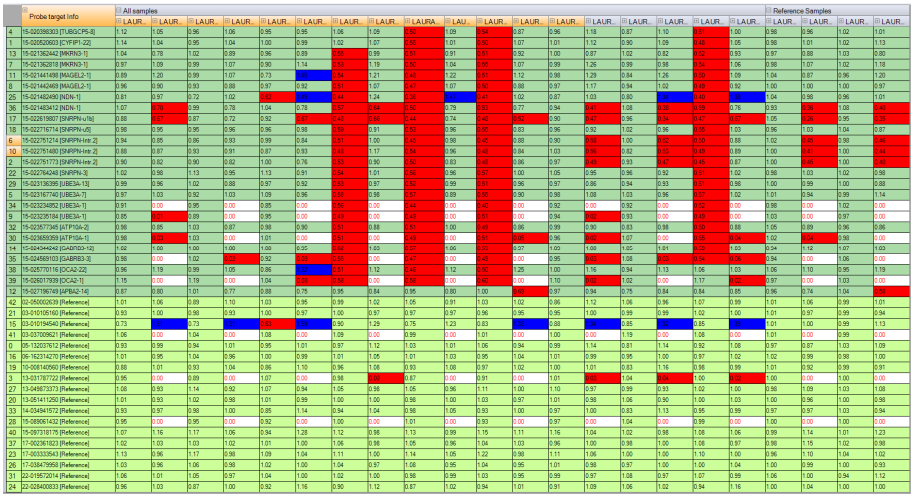


Figure 52 *Comparative analysis sample report grid on dual modus. DNA and MS-MLPA results are organized next to each other.*

9.4 9.4 Comparative analysis sample explorer with MS-Data

After the analysis is finished you can open the comparative analysis sample explorer as earlier described in chapter 8.4. We recommend viewing both DNA-MLPA results and MS-MLPA together; to make this process easier results of coupled samples (undigested to digested) are listed right underneath each other. Each normalized digested methylation sample result will be placed directly under the DNA-MLPA results with an added tag “[d]” (figure 53).

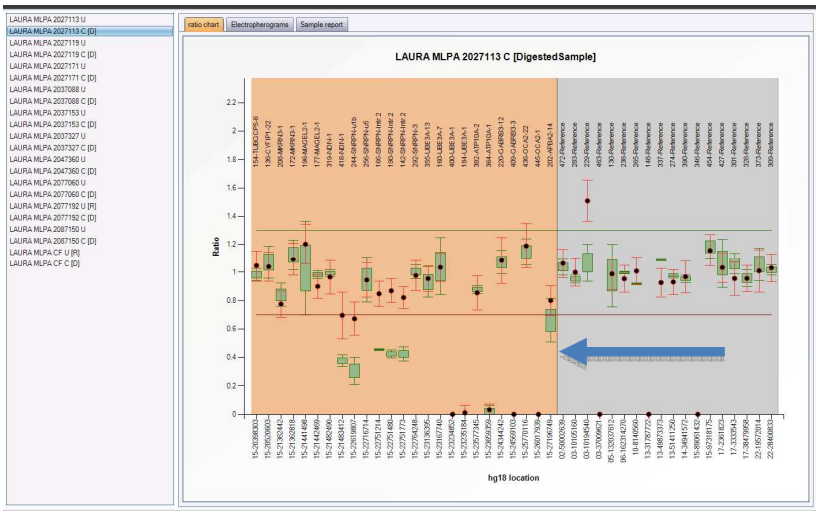


Figure 53 *Sample explorer result chart of digested reference sample.*

Different from the DNA-MLPA analysis digested samples are not normalized against the set reference samples. The reference sample are however used to create distribution statistics thereby having something to compare to, a so-called “normal situation”. In figure 54 you may for instance recognize the earlier described reference box plots (chapter 8.4.1), which in this case do not fluctuate around 1. This is the result of the normal methylation status of the SNRPN gene. A genomic imprint causes the maternal copy of this gene to be always methylated while the paternal copy remains unmethylated. The reference sample population box plot of the methylation status for this gene therefore falls around 50%, since the signals of the paternal copy are cut away in the digested sample as opposed to the undigested sample (blue arrow, figure 54). Signals that are higher than this distribution box and are higher than 75% or ratio 0.75, as opposed to the undigested sample are expected to originate from 2 uncut copies, or two methylated copies (in case this sample was found to be diploid for target sequence). In case of a Prader Willi syndrome this would most likely be caused by a uniparental disomy, where the 2 copies of this gene are received from the maternal DNA, and thus both being methylated (figure 53 & 54). In case we found by the DNA-MLPA analysis that this sample only has a single copy for this target sequence, then this copy will also be methylated still resulting in two non-functional copies (figure 55 & 56). From this we thus may deduct that it is always necessary to evaluate the methylation status in combination with the copy number status. When analyzing tumor samples situation may be even

more complex, samples that have target sequences that are triploid may for instance have a methylation percentage of 0%, 33%, 66% or 100% coming from respectively 3 methylated copies, 2 methylated copies, 1 methylated copy or no methylated copies at all.

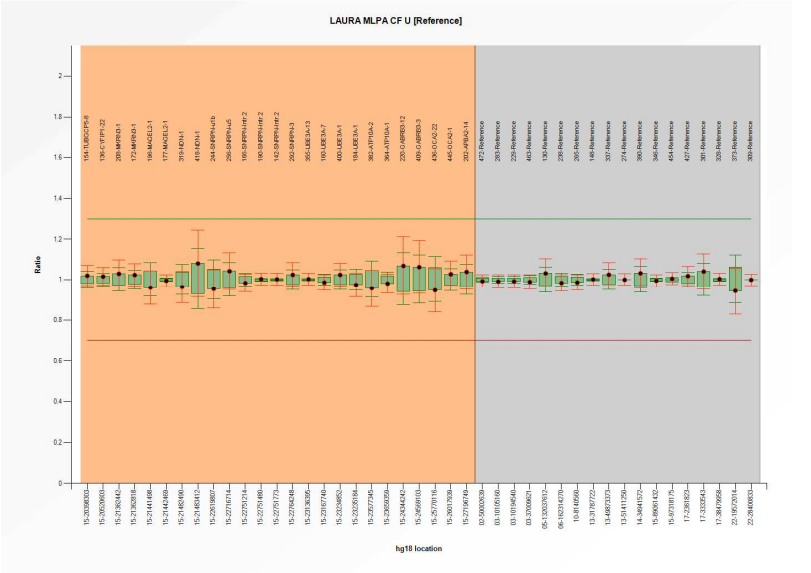


Figure 54 *Sample explorer result chart of undigested reference sample (matched undigested sample to the result of figure 53).*

Title	Coffalyser.NET analysis manual beta version
Status	Release candidate
Classification	Confidential
Versie	0.1

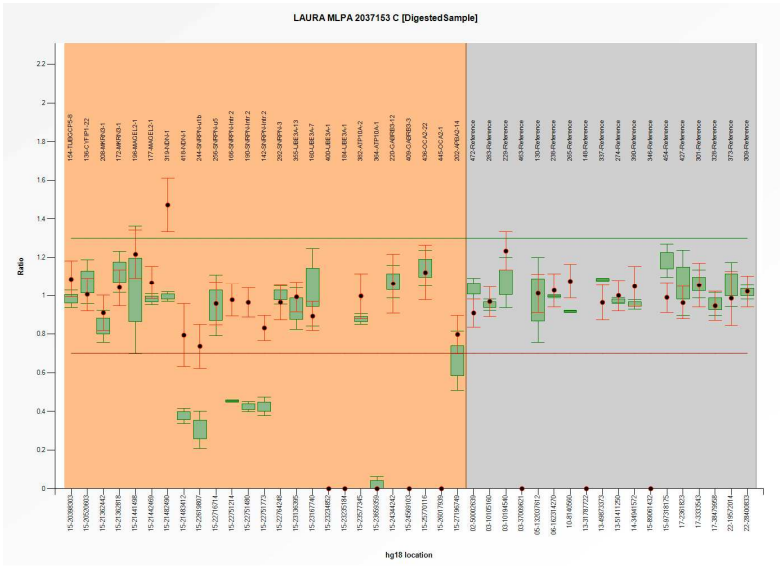


Figure 55 *Sample explorer result chart of digested sample showing a complete methylation of all target sequences except for the digestion controls.*

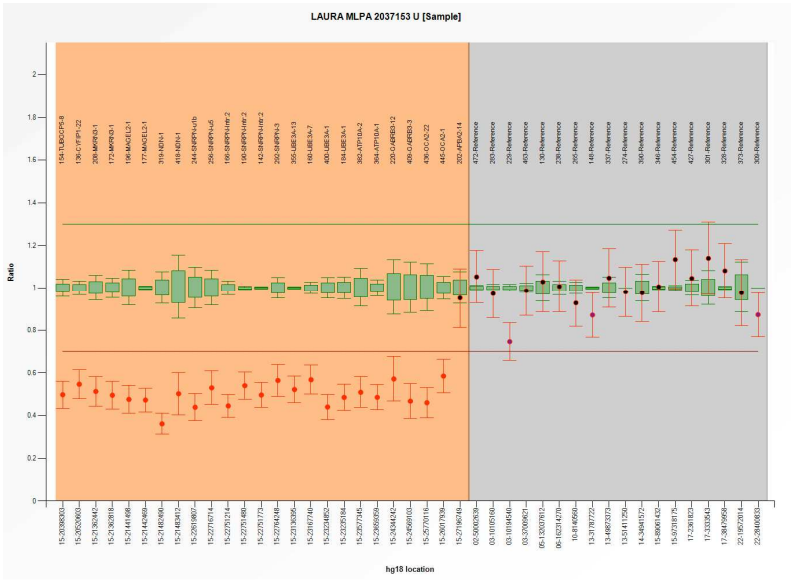


Figure 56 *Sample explorer result chart of undigested sample showing a deletion of all test probes (matched undigested sample to the result of figure 55).'*

Title	Coffalyser.NET analysis manual beta version
Status	Release candidate
Classification	Confidential
Versie	0.1

10. FAQ

10.1 What is the different when the analysis method is set to RNA

RNA-MLPA is very similar to the DNA-MLPA analysis method except that the used normalization factor will always be comprised out of the reference probes. Next to this slope correction methods are not allowed since the probe signals can never define the amount of sloping on signals originate from such different numbers of targets sequences as is found with RNA sequences. You may however set reference samples, which in case of RNA serve for instance as a zero time point while all samples are measurements on later time points. In case you wish to only investigate the intra-normalized signals against a reference (e.g. B2M), you need to adjust the Y-values or probe ratios to the intra-normalized ratios in the results screens.

11. References

Most information has been directly copied from the following book chapter which can be viewed freely online.
Coffa, J. (2011). Analysis of MLPA data using novel software by MRC-Holland, Coffalyser.NET, Intech open acces publishing, http://www.intechweb.org

1) Ahn, J.W. (2007). Detection of subtelomere imbalance using MLPA: validation, development of an analysis protocol, and application in a diagnostic centre, BMC Medical Genetics, 8:9
2) Albert, J. (2007) Bayesian Computation with R. Springer, New York
3) Applied Biosystems. (1988). AmpFℓSTR® Profiler Plus™ PCR Amplification Kit user's manual.
4) Bickel, Peter J.; Doksum, et al. (2001). Mathematical statistics: Basic and selected topics. **1**
5) Clark, J. M. (1988). Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res* 16 (20): 9677–86.
6) Coffa, J. (2008). MLPAnalyzer: data analysis tool for reliable automated normalization of MLPA fragment data, Cellular oncology, 30(4): 323-35
7) Ellis, Paul D. (2010). The Essential Guide to Effect Sizes: An Introduction to Statistical Power, Meta-Analysis and the Interpretation of Research Results. United Kingdom: Cambridge University Press.
8) Elizatbeth van Pelt-Verkuil, Alex Van Belkum, John P. Hays (2008). Principles and technical aspects of PCR amplification.
9) González J. 2008. Probe-specific mixed model approach to detect copy number differences using multiplex ligation dependent probe amplification (MLPA), BMC bioinformatics, 9:261
10) Hermsen M., Postma C. (2002). Colorectal adenoma to carcinoma progression follows multiple pathways of chromosomal instability, Gastroenterology, 123 (1109-1119)
11) Holtzman NA, Murphy PD, Watson MS, Barr PA (1997). "Predictive genetic testing: from basic research to clinical practice". *Science (journal)* **278** (5338): 602–5.
12) Huang, C.H., Chang, Y.Y., Chen, C.H., Kuo, Y.S., Hwu, W.L., Gerdes, T. and Ko, T.M. (2007). Copy number analysis of survival motor neuron genes by multiplex ligation-dependent probe amplification. *Genet Med*. 4, 241-248.

13) Janssen, B., Hartmann, C., Scholz, V., Jauch, A. and Zschocke, J. (2005). MLPA analysis for the detection of deletions, duplications and complex rearrangements in the dystrophin gene: potential and pitfalls. *Neurogenetics*. 1, 29-35.
14) Kluwe, L., Nygren, A.O., Errami, A., Heinrich, B., Matthies, C., Tatagiba, M. and Mautner, V. (2005). Screening for large mutations of the NF2 gene. *Genes Chromosomes Cancer*. 42, 384-391.
15) Michils, G., Tejpar, S., Thoelen, R., van Cutsem, E., Vermeesch, J.R., Fryns, J.P., Legius, E. and Matthijs, G. (2005). Large deletions of the APC gene in 15% of mutation-negative patients with classical polyposis (FAP): a Belgian study. *Hum Mutat.* 2, 125-34.
16) Nakagawa, Shinichi; Cuthill, Innes C (2007). "Effect size, confidence interval and statistical significance: a practical guide for biologists". *Biological Reviews Cambridge Philosophical Society* **82** (4): 591–605
17) "NCBI: Genes and Disease". NIH: National Center for Biotechnology Information (2008).
18) Redeker, E.J., de Visser, A.S., Bergen, A.A. and Mannens, M.M. (2008). Multiplex ligation-dependent probe amplification (MLPA) enhances the molecular diagnosis of aniridia and related disorders. *Mol Vis.* 14, 836-840.
19) Schouten, J.P. (2002), Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Research, 20 (12):e57
20) Scott, R.H., Douglas, J., Baskcomb, L., Nygren, A.O., Birch, J.M., Cole, T.R., Cormier-Daire, V., Eastwood, D.M., Garcia-Minaur, S., Lupunzina, P., Tatton-Brown, K., Bliek, J., Maher, E.R. and Rahman, N. (2008). Methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA) robustly detects and distinguishes 11p15 abnormalities associated with overgrowth and growth retardation. *J Med Genet.* 45, 106-13.
21) Sequeiros, Jorge; Guimarães, Bárbara (2008). Definitions of Genetic Testing EuroGentest Network of Excellence Project.
22) Taylor, C.F., Charlton, R.S., Burn, J., Sheridan, E. and Taylor, GR. (2003). Genomic deletions in MSH2 or MLH1 are a frequent cause of hereditary non-polyposis colorectal cancer: identification of novel and recurrent deletions by MLPA. *Hum Mutat.* 6, 428-33.
23) Wilkinson, Leland; APA Task Force on Statistical Inference (1999). "Statistical methods in psychology journals: Guidelines and explanations". *American Psychologist* **54**: 594–604. doi:10.1037/0003-066X.54.8.594.
24) Yau SC, Bobrow M, Mathew CG, Abbs SJ (1996). "Accurate diagnosis of carriers of deletions and duplications in Duchenne/Becker muscular

dystrophy by fluorescent dosage analysis". *J. Med. Genet.* **33** (7): 550–558. doi:10.1136/jmg.33.7.550.
25) Zar, J.H. (1984) *Biostatistical Analysis.* Prentice Hall International, New Jersey. pp 43–45

12. Appendixes

Criteria for each machine for quality control checks